

1091

~~11168~~

2/26





*Social
Psychology
through
Experiment*

7/22

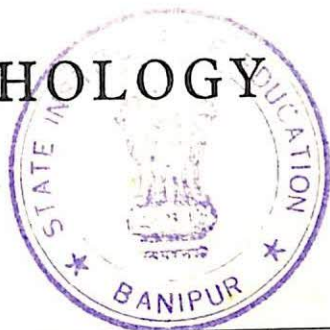


JOHN F. W. LIBRARY

Date 13. 2. 2002

Access No. 10386

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT



MICHAEL ARGYLE, *Oxford University*

GUSTAV JAHODA, *Glasgow University*

J. F. MORRIS, *Manchester University*

A. N. OPPENHEIM, *London School of Economics*

D. R. PRICE-WILLIAMS, *London School of Economics*

PETER ROBINSON, *Hull University*

HENRI TAJFEL, *Oxford University*

THELMA VENESS, *Birkbeck College*

Edited, with Introductions,

by GEORGE HUMPHREY and MICHAEL ARGYLE



BOMBAY • CALCUTTA • NEW DELHI
MADRAS • LONDON • NEW YORK
ASIA PUBLISHING HOUSE

First published in 1962
© 1962 by Methuen & Co Ltd
Printed in Great Britain

JOHN W. B. LIBRARY

DATE

ACQ. No.

CONTENTS

Preface	<i>page vii</i>
Introduction for Humanists	I
Introduction for Psychologists	II
1 Social Perception	20
Social factors in perception	22
Labels and people	32
Inferring personality traits	42
2 Motivation and Conflict	55
Achievement and level of aspiration	58
Sexual motivation and conflict	66
Super-ego conflict and guilt	71
3 Communication	79
A study of rumour	80
Content analysis	86
Stereotypes	96
4 Small Social Groups	105
Co-operation and competition	109
Communication paths	115
Group size and problem-solving	123
5 Interviews	131
The re-interview factor	144
The sex factor	148
The wording factor	150

CONTENTS

6 Social Class	<i>page</i> 155
Social class and perception	156
Friendship choices at school	163
Child-rearing practices	171
7 Small-scale Social Surveys	181
Index	204

PREFACE

The aim of this book is to provide a handbook and guide to practical class work in social psychology. In the teaching of scientific subjects the laboratory work is always an important part of the course, but social psychology is often taught without any practical work at all – largely because of the inherent difficulties. In the present volume, seven topics central to the subject matter of social psychology have been selected, and for each one an expert has been invited to describe three good experiments suitable for practical class work. Each experiment can be completed in a two-hour session; most of them yield results which are not obvious to common sense, and which are of some fundamental importance; in every case statistically significant results can be obtained under the limited conditions described. It is hoped that this book will stimulate practical class work in this subject and that it will serve as a useful guide to teachers and students alike. The editors have dared to hope also that non-psychologists, who may perhaps be startled by the idea of experiment in this field, may be induced to look over this volume in order to see the kind of thing that is being done. For that reason the senior editor has reprinted from a companion volume *Psychology Through Experiment* (Methuen) part of the 'Introduction for Humanists' (slightly modified), which was intended to serve the same purpose. *The Editors will welcome communications from interested readers, professional or otherwise; practical hints concerning the conduct of experiments will be especially welcome.*

G. H.
M. A.

THE HISTORY OF THE

REIGN OF

CHARLES THE FIRST

IN WHICH ARE CONTAINED THE
CAUSES, THE PROGRESS, AND THE
CONSEQUENCES OF THE
GREAT SCALES

OF THE
COMMONWEALTH
OF GREAT BRITAIN
IN THE
SEVENTEENTH CENTURY
BY
JOHN BURNET
OF THE
UNIVERSITY OF OXFORD
IN TWO VOLUMES
THE FIRST

CONTAINING
THE
LIFE OF
CHARLES THE FIRST
FROM HIS BIRTH
TO HIS DEATH
IN THE YEAR
1649

INTRODUCTION FOR HUMANISTS

GEORGE HUMPHREY

Surprisingly often the question is asked why psychologists do experiments. On both sides of the Atlantic it is the humanist who seems particularly disturbed by our experimental activities. The laboratory, the animal house and the slide rule are often faintly repulsive to him and he feels that the last sanctuary of the human spirit is being invaded.

Two cases. Two of many current cases will illustrate the point. A philosopher of deservedly international reputation asked not long ago why psychologists bothered any more about instinct, when 'All you have to do is to look in McDougall's book [meaning his remarkable *Social Psychology*, first published in 1908] where you will find it all done up nicely. He's settled the question once and for all.' It was indeed done up nicely – too nicely, admirers of the experimentalist Tinbergen's *The Study of Instinct* (1951) will think.

Then in general there is the distaste many highly cultured people feel at the thought of experiment on human minds. To them there are delicacies which must be respected, even in these days of governmental controls and of living in houses of civil-service glass, and this is one of them. 'There is,' somebody has said, 'an unseemly exposure of the mind as well as of the body.' The social scientist is:

*a fingering slave
one that would peep and botanize
upon his mother's grave,*

asking a man as he does how he is going to vote (which is emphatically a man's own business), or whether he would want his daughter to marry a negro or an eskimo (which is his daughter's). And so one sees a national newspaper suggesting, only half seriously indeed,

that WON'T ANSWER should be included as a heading in public opinion polls, alternative to YES and NO. An Englishman's Home, it is felt, is his Castle; his Mind is his Holy Place, where strangers must not tread. 'Do you mean to say,' asked a personal friend in tones of hushed outrage, 'that if you were psychoanalysed you would tell *everything* that came into your mind?' and last evening (September, 1959) a highly competent young philosopher said: 'I don't *want* you people scrutinizing me.' Such horrified reaction against the invasion of personal privacy is of course very proper in our modern society, and there is no doubt that there have been times and occasions where it has seemed to many humane people to spread contrary to the interests of society as a whole. Browning is said to have flung himself off to Italy in a dudgeon because of the outrageous questions about his private and personal income the tax officials asked him. But when slaves were freed in British lands, the government compensated slave-holders for the loss of their 'property' in flesh and blood, and most good people seem to have thought this was entirely just. 'Just' it most certainly was. Yet we now shudder at the total implications of this piece of righteous compensation.

Tempora mutantur et nos mutamur in illis. Truly the last hundred years or so have seen the final abolition of the traffic in human beings, the institution of 'planning' authority, the Married Women's Property Act and other innovations that seemed to many virtuous people of the time infringements of human personality and human rights. Many things once defended as inalienable privacies are yielding to public opinion and public good. Experiment in psychology is part of the swelling tide. Social psychology is part of the defence against improper invasion of the spirit; for it aims to understand what is happening, and understanding leads on to control and defence. It is today not enough for a man, however clever, to watch the behaviour of other people as he goes about his business and write a book about the Social Contract or what have you. We have suffered too much in the past from this sort of thing. The powerful modern method of experiment must be used to take the uncertainty from the answers. If the techniques may seem at first to savour of some impertinence, the social psychologist must, rather grimly, allow him-

self to be tried in the same dock with William Pitt the Younger, who invented Income Tax.

Why experiment at all?

To experiment is roughly to check one's opinion of Nature against Nature herself. The foundation of much humanistic education lies in books; perhaps for this reason humanists in all ages, as a class, have been against experiment as an educational method. It seems to many highly cultured people to consist of 'fiddling with bunsen burners and turning knobs', not to involve 'real thought', and to be a sorry substitute for 'the study of the great minds'. But it will be seen that in point of fact books have often done great harm by crystallizing and ultimately perverting opinion, and that far from concerning itself with trivialities, scientific training and thinking follow a uniquely difficult path of intellectual humility. This intense kind of work necessarily comes into good experimental teaching, just as the benefits of 'the classics' issue from their own kind of teaching when it is well done – but only then.

Before dealing with the particular case of social psychology the general problem will first be considered. Why experiment at all? Artists go out to sketch and often use human models. In much of their training, also, human models are employed. Scientists go to nature in the same way . . . physical scientists to inanimate nature, physiologists and psychologists to living creatures. Carlyle once wrote: 'The place where we are to get knowledge, even theoretic knowledge, is in the Books themselves.' This is valiantly spoken; as a universal and unqualified rule however it is just not true. Newman tells us, for example, that in Athens at its heyday there were no bookshops at all. Of course the scientist has to go to books, but finding out what people have written is only a part of it. He is interested rather in what they have done. For reading is a secondhand way of holding the mirror to nature. As Einstein has said: 'If you want to find out anything from the theoretical physicists about the methods they use I advise you to stick closely to one principle: don't listen to their words, fix your attention to their deeds' (quoted by Toulmin, 1953, p. 16). And in the fifth century of this era Proclus

the Neoplatonist philosopher tells of a man who found in the Greek countryside a lizard showing spots that were not according to Aristotle. His conclusion was that the lizard was wrong. When again Galileo di Galilei's astronomical reasoning and observation seemed not to conform with what was written, it is well known that a sermon was preached against him on the text: 'Ye men of Galilee why stand ye gazing up into heaven?' though it is fair to say that Galileo received an apology from the head of the order involved. His contemporary Scheiner who claimed priority over him for the discovery of sun-spots was told, on communicating what he had seen: 'I have read Aristotle's writings from end to end many times, and I can assure you I have nowhere found anything similar to what you describe. Go, my son, and tranquillize yourself; be assured that what you take for spots on the sun are the faults of your glasses, or of your eyes' (Fahie, 1903, p. 130). So far can books tyrannize over facts, and so far must Carlyle be allowed to be wrong. On the other side of the scale one may see an undergraduate, being shown something he has read about, not uncommonly realize with surprise that 'it really works!' What surprises him is that what the books tell him really happens *to him*, so to speak. Before, he had 'book knowledge'; now he has 'personal knowledge', and the exclamation is the mark of the transition from one to the other.

Humanists against experiment. It is probable that the humanist has fought shy of experiment in all ages and for much the same reasons, though these are perhaps not always explicitly stated or even realized. Teaching men to control men has seemed a higher aim than teaching them to work with their hands. Skill in the use of military arms has always been an exception. The reason the ancient Greeks made little progress in experimental science is said to be that they felt manual work was somehow degrading for a free man. The Athenian man thought it comic and mildly unworthy that people should interest themselves in commonplace things such as the buzz of a mosquito (Aristophanes' *Clouds*), though in view of the alleged prevalence of malaria in the ancient world a little knowledge of this insect would not have been a dangerous thing. A trace of the same attitude seems to linger somewhere in this still persistent attitude of the

humanist. Experiment is often disturbing or brings results disturbing to people of fixed – and irreproachably honest – opinions. To those who have never practised it, and therefore do not see it in its context, it does seem trivial, fiddling and somehow not worthy of the attention of the best minds and the best men, who ought, it is felt, to be occupied better than by peering into test-tubes and in general playing peeping-Tom into trivialities. Not so long ago, and in the presence of the writer, a philosopher laughed uproariously when told that a student had obtained a doctorate with a thesis describing in new detail the processes at work when an earthworm moves, an activity of great interest to biologists and one which forms a paradigm to important human functions. But wisdom is justified of all her children. Experiment often leads to material progress, which itself is often disturbing. Classroom experiment is indeed often trivial and fiddling and often accompanied at school by tomfoolery; but not more so than learning the basis of any technique, from irregular verbs to five-finger exercises, and – that old example – the multiplication table, whether one actually learns the table or talks about matches in a matchbox. And the fiddling with test-tubes, the turning of knobs and other such manual activities, together with the laborious computing of probabilities, are of course in well-designed laboratory work as well as in research an insignificant-seeming though necessary part of the total experimental activity, which may call for an intellectual power, a concentration and a patience at any corresponding stage as great as that which any humanist may practise. Would anybody look down on Shakespeare because he once spent a large part of his time fiddling with a goose-feather? The parallel only seems trifling because we take the technique of writing so much for granted, forgetting the infinitely laborious preoccupation with trivialities by which we all had to learn it, or even that these small movements, acquired with or without tears, are trivialities only when considered out of their context. The time will come when those responsible for the teaching of the succeeding generation will regard learning to experiment as being equally important with learning to read and write, and will realize that being taught to light a bunsen burner or to read the level of the liquid in a graduated tube is as

fundamental as learning any of the basic skills of the primary classroom of today. It was not so long ago in the history of our race that to possess the everyday skills of our society was regarded as unworthy of a noble man, or even, farther back, of a free one.

Bronowski. It is the same kind of thinking. In a recent discussion of the relation of science and society, Professor Bronowski made the point that the historian and the experimentalist use the same kind of reasoning process, and that as far as their thinking goes there is little difference between the two, or between the humanist and the scientist in general. Of course he is right. Human beings use the same basic mental processes whatever problems they are engaged on. But the experimentalist has this great advantage that he is able to check his conclusions by the *appearance* of nature, at least, not merely by the analogy (Bronowski) of other human beings. This direct comparison of hypothesis with fact is an advantage the historian would be glad to borrow. 'What Song the Syrens sang' we are told he can only conjecture; and however close his personal knowledge of songs and singers, he would be glad enough to check up on his thinking, to try the experiment and himself listen to the charmers – or for that matter to tape-record their performance.

The battle is not yet won. It will be said at this point that experiment and the teaching of experiment in general have long been accepted by humanists and that what has been said is shouting after the war is over. It is true that the Royal Society was founded over three hundred years ago, for the purpose among other things of practising 'the new experimental philosophy'; and that its activities and those of its members have won respect in Great Britain and everywhere else; further that largely under its leadership and that of similar organizations throughout the world the growth of experimental science during the last three centuries has everywhere been enormous. But some may be surprised to hear that in 1869 Helmholtz, who died only six years before this century began, was refused the chair of Physics at Oxford, now officially so liberal in these matters. According to Cyril Darlington, 'an agreeable young . . . mathematician . . . with a considerable estate in Lincolnshire' was instead appointed by the five electors, of whom three were scientists

'certainly not interested in experiment' . . . The new professor 'lived to a great age and for just fifty years [i.e. until after the First War] he was successful in forbidding all new physical experiments in the Clarendon Laboratory'. Helmholtz on the other hand 'proved to be one of the greatest influences in the development of science, and his influence stretched beyond his own country and beyond his own time' (1958, quoting *Nature*, 1921). *It is simply not true that the humanist – speaking generally – has accepted experiment on completely equal terms with his own kind of study.* Though the story is probably fabled of the classical headmaster who, when his chemistry specialist was taken ill just before an examination had to be set, remarked that he must 'take a day off and get up the subject'. But it is true that many humanists find it hard to believe that experiment demands real thinking.

Jargon. Finally, to anticipate, psychologists with other scientists have been reproached in high quarters for using 'jargon'. (Not so long ago they were reproached for the opposite fault, of 'giving bad reasons for what everybody knows already'.) It is hardly necessary to point out that for precision of thought new terms may have to be invented. The writer would also remind humanists that when he was being taught to understand Vergil it was considered necessary for him to know and use the terms *asyndeton*, *aposiopesis*, *hemistich*, *zeugma*, *tnesis*, and *chiasma*.

Why think? Try the experiment. The great Jenner, discoverer of vaccination, is often quoted. He is said to have asked the question at the beginning of this paragraph, but of course he did not mean that experiment and thought are mutually exclusive. The way he made his great discovery shows this. He first observed that milkmaids who had cowpox did not contract smallpox, and then had the thought that other people also might be protected by infecting them with the less serious disease. That is, he formed a hypothesis by a process of thinking. A hypothesis may be called an expendable opinion. But instead of continuing to speculate he thereafter tested the hypothesis by trying the experiment. It is well for mankind that he did.

The scientist and the model. Lord Kelvin used to say that he could not think properly about any physical event until he had visualized

a mechanical model of it. The term was extended by another great physicist, Mach, Kelvin's contemporary, who claimed that the atomic theory . . . is a mathematical *model* for facilitating the mental reproduction of facts. (Mach's italics. See his *Principles of Mechanics*, Open Court, Chicago, 1893, p. 492.) The implication intended by Mach was that such terms cannot cover the whole of nature but only certain relations in it that can profitably be described, ideally in mathematical terms. The term has spread until it is now used to describe a mental construction, derived from observation, with which certain aspects of a natural event conform, and which can be used to suggest further experiment. Thus what used to be called the association theory in psychology could now be said to serve as a rough model for certain mental events; one idea revives a second one which has been related to it in one of several specified ways. This statement was thought to be derived from experience; it was used in fact as the basis of many experiments; in an extended form it was used by the conditioned reflex workers for further experiments, and was put into mathematical terms by American workers and others. But it does only describe certain aspects of the events with which it deals. One of Pavlov's dogs secreted $13\frac{1}{2}$ drops of saliva at the sound of a metronome previously associated with food. We are told in Pavlov's *major* work the time and the date, but not the length of the dog's tail, or the colour of the experimenter's hair. The book quite properly abstracted from all these things as 'irrelevant'. Every natural event is greater than description can ever be; abstraction must always be made, as the artist and the scientist both know. And of course, this does not mean that the models of science (or the pictures of the artist) can be *deduced* from natural events by the ordinary processes of logic. Einstein has pointed this out (Toulmin, 1958, p. 43). Genius knows what to abstract *from*.*

The scientist and the way of humility. The way of experiment is on all sides the way of humility, of the man who knows he can never know all of reality, but can only deal with a small part of it; who checks his hypotheses, *wherever he obtains them*, against nature

* In the special theory of relativity abstraction was from the observer's point of view.

herself; whose opinions, to repeat, are expendable. A good experimentalist cannot be a 'know-all', for he knows too many difficulties. If he has been well taught he sees problems where others see dogmas. To know what must be accepted and what should be questioned involves intellectual activity of the highest order. 'He knows nothing and thinks that he knows,' said Socrates. 'I am better off, for I neither know nor think that I know.' The path of humility as taught by Socrates may serve as a motto for all entrusted with the task of teaching, whether primary or university teaching, experimental or humanistic. It is a humbler way than that of the man 'prepared to stand by his opinions', wherever acquired, from whatever authority and however 'ethical' or unethical in character; or boasts, as some do, that he knows 'everything that has been written' on this or that. Properly carried out it requires generosity of spirit and tough, disillusioning thinking. It is a fair rule that the better the scientist, the humbler the man – at least about his subject.

Newton wrote: 'I do not know what I may appear to the world, but to myself I seem to have been only a boy playing on the sea-shore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, while the great ocean of truth lay all undiscovered before me.' Of him, in the year of his death, Pope wrote: 'God said *Let Newton be* and all was light.' The quotations are apt to be remembered separately, and are generally forgotten together.

This then is a plea for experiment in general both as involving an intellectual activity of the highest power, and as a pre-eminently searching method of education.

Experimental thinking often makes dangerous thoughts. A final word: especially in view of the growing social importance of Science it is today of transcendent urgency for the future of mankind that the coming generation shall be taught how Nature is thus interrogated, human and otherwise. If axioms, apparently self-evident, have to be questioned, and they must be, everybody must know why and how. It is at its peril that mankind will listen without questioning to the voice of that philosopher who has told us that whatever is clear and distinct is true. What could be more clear and distinct than that

the sun moves every day across a stable sky? Or that Nature abhors a vacuum, when one can hear and see Nature abhorring a vacuum every time water is drawn in a pump? In spite of the air of slightly discreditable nonconformity and even of immorality that such enquiries often involve, such questions must be asked and people must be taught to ask them, or at least how they are asked. Few of course can learn to experiment like a Nobel prizeman; but even fewer learn to write like Shakespeare, in spite of the millions of hours still being spent on the drudgery of handwriting and literary composition.

When any branch of knowledge can experiment it must experiment or it will die a dreary death.

What of Social Psychology in particular?

REFERENCES

FAHIE, J. J. (1903). *Galileo; his life and work*. London.

TOULMIN, S. (1958). *The Philosophy of Science*. London: Hutchinson.

INTRODUCTION FOR PSYCHOLOGISTS

MICHAEL ARGYLE

Problems with practical class work in the teaching of social psychology. It is generally agreed that scientific subjects cannot be taught properly unless the student can have first-hand experience of the phenomena being described. In one sense everyone has this experience already in the case of social psychology, but no one has experience of the special research techniques used, or of the precise results which are obtainable under controlled conditions. Unless people actually have some experience of systematic observation, scoring projection tests, and designing questionnaires, they cannot appreciate the scope and limitations of these methods. Unless they actually see some of the classical results demonstrated they will never really believe them. Unless they tackle a new experimental problem and test a new hypothesis they will never get the feeling of psychological research.

For years now, students of experimental psychology have tried out in practical classes the traditional experiments on perception, non-sense-syllable learning, reaction times and so on. Many of these experiments can be done conveniently in an afternoon with one subject, so that students work in pairs and take turns to act as subject. The most obvious difficulty with practical work in social psychology is the greater number of subjects needed. A typical small group experiment reported in the literature might involve eighty groups of five people—400 subjects in all; a survey of attitudes would require from 500 to 3,000 respondents, and so on. Considerable ingenuity is needed to get round this problem. With small group experiments, the members of the class can be the subjects, and there might perhaps be twenty of them. As Thelma Veness describes in Chapter 4, groups of three may be used, and in other cases groups of two. If the same groups are used in two experimental

conditions, twenty subjects provide ten matched pairs of groups, which are sufficient to give significant results.

In social surveys, as Morris describes in Chapter 7, 100-150 respondents can easily be found in the course of an afternoon by twenty student-interviewers, and this makes a quite adequate sample for a pilot survey, from which very interesting results are often obtained. For other topics, such as perception and motivation, the individual members of the class provide enough subjects. With this kind of experiment, three experimental batches of five subjects can often give highly significant results, if one uses a non-parametric statistical test.

The problem of subjects can be met, then, either by the use of all the members of the class, or by the use of outside subjects. If the students are to be the subjects, they should also be given a turn as experimenters - several are needed in some of the experiments - and of course all can participate in working out the results. Outside subjects can be employed in a variety of ways. In a social survey, names of other students can be chosen to form a stratified sample, and several names given to each interviewer to find. In a laboratory study of interviewing, as described by Price-Williams in Chapter 5, volunteers can be obtained, from another psychology class for example, who will come to the lab at a certain time. In a study of child-rearing, such as the one described by Jahoda in Chapter 6, it may be necessary to approach a number of mothers beforehand - those known to the person in charge for instance - to explain the project and ensure co-operation. If a research contact exists with schools, factories or prisons, the students may be able to take part in the research in some way, though this will need skilful handling in view of the delicacy of some of these social situations.

A second general problem with practical class work in social psychology is that none of the experiments can be done without the exercise of special skills and techniques. The solution to this problem is to make training in these techniques a definite part of the course. In the experiments on Communication, described by Oppenheim in Chapter 3, the students will have to learn something about content

analysis; to do the experiments on interviewing in Chapter 5 they will have to learn how it is done, and they will need this before engaging in the social survey. They will also learn methods of observing small group behaviour (Chapter 4), and it is well worth trying out simple methods of interaction recording for their own sake, without making any experimental use of them. They will learn how to score projection tests (Chapter 2) and how to design samples and questionnaires (Chapters 6 and 7). Another skill which is most important in this type of investigation is the general handling of, and demeanour towards, subjects: this can partly be done by example and partly by explicit instructions, while role-playing may also be useful in this connection.

In addition to the teaching of special research techniques, some explanation of statistical methods may be necessary, unless they have already been covered. In any case there will have to be some discussion of *which* statistical tests are appropriate. The writers in this book, following current practice, have recommended non-parametric methods, which are more suitable to most of the data of social psychology than are the traditional methods. The book by Siegel (1956) gives an admirably clear account of these newer methods, and could be used by students in conjunction with these investigations.

A problem is created by the use of members of the class as subjects – they may already be familiar with the experiment and its anticipated results. For example, those who have read Atkinson's volume on Motivation (1958) will easily recognize attempts to arouse achievement motivation. However, there are numerous variations on this type of experiment, and the student may not realize if it is in fact aggression which is the object to the experiment (see Chapter 2).

It is a very good thing in this sort of work to try out new hypotheses all the time; viewed in this way, the experiments in this book can be used as a guide to the sort of things that are feasible in the classroom. It is probably best to include at least one hypothesis that can be relied upon to work, together with new ones devised by the class or the person in charge. This gives students the feeling of

research, convinces sceptics that common sense cannot explain all, and may lead to something actually being found out!

The case for applying experimental methods to the study of social behaviour. Many people think that social behaviour should not be studied by systematic or experimental methods at all. They feel that the dramatist and novelist, the historian, and perhaps the psychoanalyst, can do more by means of clever intuition and observation. While it is true that we need ideas and hypotheses from all of these and other sources, there is only one means of finding out which hypothesis is correct and under what conditions – that is by systematic experimental and allied research. It is the aim of this research in the first place to build up a solid body of empirical results, so that behaviour can be predicted from its antecedents, and then to develop theories which will explain these empirical laws.

It has been doubted whether human social behaviour is sufficiently predictable in principle for such laws to be found. Many examples could be given to show how very predictable human behaviour turns out to be. For instance the Gluecks (1950) constructed a prediction table for distinguishing between boys who would and would not become delinquent, on the basis of their early relations with their parents. Those scoring above a certain point in the table have a higher chance of becoming delinquent, those below another point have a much smaller chance.*

It has been objected that the results of much social research are not surprising, that we knew them all before. There is an easy way of convincing people that this is often not so. Present an experiment from the literature in full detail, but without the results: ask an audience to predict what they will be. A number of quite different suggestions will be put forward, and if the experiment is carefully chosen, all of them will be wrong! There is nevertheless a kernel of truth in the complaint, since investigations are sometimes carried out whose conclusions were never really in doubt, though it may be useful to have them empirically confirmed. This would be avoided if investigations were only undertaken when there is genuine doubt concerning the outcome. If an hypothesis is almost certainly true there

* A valuable critique of this work is provided by Stott (1960).

is little point in confirming it, if almost certainly false it will probably be a waste of time testing it, but where the subjective probability is about half, maximum information is obtained by testing it.

The best answer to this objection is to give examples of results which were unexpected. A good source of surprising results is psychoanalytic theory. An area in which Freud's ideas are fairly well confirmed is that of fantasy. The second experiment in Chapter 2 of this book shows how sexual needs may be expressed in fantasy – by telling stories about TAT cards; where conditions of restraint are present this indirect form of release will take place, including the use of Freudian symbolism. An experiment by Feshbach (1955) showed that such expressions in fantasy were successful in satisfying the drive.* An area in which Freud's ideas have on the whole been refuted is that of the relation between early childhood experiences and later personality. There is no evidence that severe toilet training produces an anal character, and there is some doubt whether the anal character exists – i.e. whether the traits of which it is composed correlate with one another (Child, 1954).

Research into social groups has produced a very extensive body of results, some more surprising than others. Festinger and Carlsmith (1959), for instance, found that subjects paid to act as confederates and tell other people something which they didn't themselves believe, changed their minds more in consequence *the less* they were paid; this gives some support to Festinger's 'cognitive dissonance' theory (1957) but otherwise is quite unexpected. Schachter (1959) found that hungry and anxious subjects preferred to wait in a room full of other subjects, compared with control subjects who were content to wait by themselves; only first-born children when anxious were particularly liable to choose the social situation. Other group examples are the three experiments given in Chapter 4 of this book.

It is sometimes objected that these experiments are artificial, that they are conducted under 'bogus' laboratory conditions, so that their results would not apply in the real world. In some areas of social psychology the investigator can choose between field studies and laboratory studies. Field studies have the advantage of being fully

* See Buss (1961) for a review of work in this area.

realistic, but have the drawback that experimental changes cannot usually be introduced: the investigator has to compare groups or individuals statistically as they are. Laboratory studies are rather less realistic, but experimental changes *can* be made, and the situation can be under full control. In this book we have concentrated on the second kind of study, partly since it is very difficult to introduce students to field situations. If laboratory experiments are used, all efforts should be made to introduce realistic conditions. However, many interesting results have been obtained under very unrealistic conditions: an example of this is Crutchfield's work on individual differences in conformity (1955), where group members did not see the other members but watched a display supposedly showing judgments made by them. Such results are of great value even if they do not precisely apply in other situations.

The drawback of the purely statistical type of field study is that it is impossible to tell which is the direction of causation. For instance, if it is found that aggressiveness in small boys correlates with the use of physical punishment by the parents (Sears, Maccoby, and Levin, 1957), this could mean that spanking causes aggression, that innately aggressive children get beaten more, or that parental aggressiveness has been inherited by the children. On the other hand it is sometimes possible to choose subjects with ingenuity so as to rule out alternatives – the use of foster children as subjects would cut out the third possibility here for example. On rare occasions it is possible to make use of administrative or technological changes as a source of experimental manipulation. Feldman (1937) studied the output of work-groups before and after the management changed all the foremen round, and discovered that the output under each foreman was much the same as before.

It is sometimes claimed that these researches do not go 'deep' enough, – not as deep for example as detailed case studies. However, while case studies may be a source of illuminating ideas and hypotheses, they cannot establish the truth of these themselves. Only by experimental or statistical research can causal relations be established, together with the extent and form of the empirical relationship, and the generality of the results. Case studies can only give us guesses

about one person or group. Again, there is perhaps a kernel of truth in the criticism – that it is very difficult, by means of rigorous procedures, to explain what is going on in particular social situations. To say something about the causal laws governing *one* person, situation or group, it is necessary to collect a large and homogeneous sample of similar cases, to establish the necessary empirical laws, and then show that the first instance is an example of these laws operating. In the study of individual personality, there have been one or two attempts to devise methods for finding laws true for individuals. Welford (1947) found correlations for one subject at a time between their likelihood of praying in six situations, and rankings of the frustration and emotional arousal of the situations. For some subjects prayer was associated with frustration, for others it went with the arousal of pleasant emotions.

It may be objected that experimental research does not make for ‘understanding’ of what is going on. The kind of understanding offered by scientific procedures lies in the deduction of empirical laws from theoretical postulates or models. In experimental psychology proper there has been a lot of successful theorizing of this type, in social psychology the position is very much less satisfactory. Most of the experiments described in this book are followed by discussion of the theoretical significance and the explanation of the results found. A number of miniature theories have been put forward, each covering small areas of findings. In addition there are two general theories which have been found very useful in social psychology – psychoanalytic theory and learning theory. As was shown above, Freudian ideas have stimulated a lot of research, and in some areas have been confirmed. At other points it has been definitely refuted, and at others superseded. As a source of hypotheses it will continue to be useful for a long time, as a theoretical system it must take its place in the past history of psychology. Learning theory, as expounded by Hull, Mowrer, N. E. Miller and others, was originally devised to account for learning in rats. It has also been found to apply to human learning in rather restricted situations such as conditioning and verbal learning. It is extremely doubtful whether it can incorporate the kind of learning which goes on during the socialization of the

child, which takes place in a social situation and is mediated by relationships with other people. It has been found by Argyle and Robinson (1962) that achievement motivation is a product of identification with achievement oriented parents, together with introjection of parental exhortations. This shows the operation of two complex forms of social learning, based on social relationships, not easily accounted for in learning theory terms.

A final objection is that the trained practitioner is sometimes said to be a better predictor of behaviour than the psychological researcher. There are a number of cases where this claim can be tested, but it should be noted that the claim can only be tested by means of orthodox research procedures. For instance, can prison governors predict failure of released prisoners better than prediction tables? Mannheim and Wilkins (1953) review a number of studies where this comparison is made. Thompson found that the Glueck prediction table mentioned above was correct in 91 per cent. of cases, whereas three experts were right in 61.5 per cent. to 65.3 per cent. of cases. Another comparison is provided by the integration of personality data for prediction purposes by a skilled clinician and by statistical methods. Meehl (1955) reviewed this material and found that the 'actuarial' method always made superior predictions. Finally, can primary school teachers make better estimates of success at grammar school than objective measures? Vernon (1957) reviews the research on this question: McClelland for example found that teachers' estimates correlated .72 with grammar school success, compared with .77 for the school exam, and .69 for an intelligence test. A combination of these three measures correlated .80 with grammar school success. This study shows that the predictions of experts can be usefully incorporated in the prediction table in this case, and that objective tests add something to those estimates.

REFERENCES

- ARGYLE, M., and ROBINSON, P. (1962). 'Two origins of achievement motivation'. *Brit. J. soc. clin. Psychol.*, 1 (in press).
- ATKINSON, J. W. (1958). *Motives in Fantasy, Action and Society*. New York: van Nostrand.

INTRODUCTION FOR PSYCHOLOGISTS

- BUSS, A. H. (1961). *The Psychology of Aggression*. New York: Wiley.
- CHILD, I. L. (1954). 'Socialisation'. Chap. 18 of G. Lindzey (ed.): *Handbook of Social Psychology*. Cambridge, Mass.: Addison-Wesley.
- CRUTCHFIELD, R. S. (1955). 'Conformity and Character'. *Amer. Psychol.*, 10, 191-8.
- FELDMAN, H. (1937). *Problems in Labor Relations*. New York: Macmillan.
- FESHBACH, G. (1955). 'The drive-reducing function of fantasy behaviour'. *J. abnorm. soc. Psychol.*, 50, 3-11.
- FESTINGER, L. (1957). *A Theory of Cognitive Dissonance*. New York: Row, Peterson.
- FESTINGER, L., and CARLSMITH, J. M. (1959). 'Cognitive consequences of forced compliance'. *J. abnorm. soc. Psychol.*, 58, 203-10.
- GLUECK, S. and E. (1950). *Unraveling Juvenile Delinquency*. Harvard Univ. Press.
- MANNHEIM, H., and WILKINS, L. T. (1955). *Prediction Methods in Relation to Borstal Training*. London: H.M.S.O.
- MEEHL, P. E. (1955). *Clinical and Statistical Prediction*. Univ. of Minnesota Press.
- SCHACHTER, S. (1959). *The Psychology of Affiliation*. Stanford: Stanford Univ. Press.
- SEARS, R. R., MACCOBY, E. E., and LEVIN, H. (1957). *Patterns of Child Rearing*. New York: Row, Peterson.
- SIEGEL, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw Hill.
- STOTT, D. H. (1960). 'A new delinquency prediction instrument using behavioural indications'. *Int. J. Soc. Psychiat.*, 6, 195-205.
- VERNON, P. E. (1957). *Secondary School Selection*. London: Methuen.
- WELFORD, A. T. (1947). 'Is religious behaviour dependent upon affect or frustration?' *J. abnorm. soc. Psychol.*, 42, 310-19.

Chapter I

SOCIAL PERCEPTION

HENRI TAJFEL

Oxford University

Social perception is not one of the best ordered areas of enquiry in psychology. Some of it is hardly 'social'; some hardly 'perceptual'. Investigations, referred to in one context or another as concerned with it, add up to a patchwork, not to a pattern. Therefore, to attempt a definition (or a series of definitions) of social perception would be a sterile task. Loosely, this blanket term stands for an agglomeration of studies attempting to find some regular and predictable relationships between man's general orientation towards his environment, and the multitude of social factors which, in one way or another, determine this orientation or contribute to it.

These relationships may be of several kinds. For example, social influences may affect our perception of the physical environment. Conversely, changes in the physical environment may affect the way in which we react to various aspects of our social environment. As any addict of *Punch* or of *New Yorker* knows, only a precarious hair-breadth distinction separates the dignified from the ridiculous: let the rain fall, or the scene be set in a tent in the jungle, and the otherwise respectable habits of dress, speech, gesture, or courtesy become fit material for a collection of 'funniest cartoons of the last fifty years'.

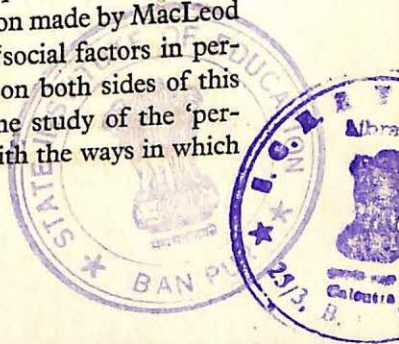
These relations between the 'physical' and the 'social' represent one only amongst the trends of interest in social perception. Our social environment consists, in the main, of people, of the ways in which they behave towards each other and towards ourselves. It is an extremely complex environment, and in order to make sense of it

we must simplify, classify and label. An individual may be of interest to us either as a member of a specified group or class of people, or in his own unique capacity. To an anti-Semite, a Jew may appear as having certain attributes: the interpretation of the Jew's appearance and actions may have its roots, entirely or predominantly, in the fact that he is known to be a Jew, and it may be to a large degree independent of the individual differences between Jews. At the other extreme, an intimate friend may be perceived or judged as being possessed of one attribute or another, independently of whether he is a Siamese, a lawyer, or a mountaineer.

It will be obvious that 'pure cases' at either extreme hardly exist. A set of characteristics may be attributed to someone because he happens to belong to a group about which we have preconceived ideas; obviously his own individual characteristics will also, under various conditions, become important. Conversely, an intimate friend may be seen as representative, in some respects, of his group; our ideas about this group will then serve the dual function of emphasizing some of our friend's traits and of providing a 'causal' explanation for their existence.

The differences between these two forms of reaching conclusions about people are quite important. Inferences based on class membership are primarily deductive; those based on individual characteristics primarily inductive. Research on the first type of inferences must be mainly concerned with the relationships that exist between the assignment of an individual to a class, and the perceptual and evaluative consequences of this assignment. On the other hand, some of the most crucial and difficult tasks in the study of inferences from individual characteristics are in discovering what cues are selected as important and relevant, what use is made of them, and what is their objective validity.

These various areas of enquiry in social perception can be grouped in a different way along the lines of the distinction made by MacLeod (1951) between 'perception of the social' and 'social factors in perception'. The points of departure of research on both sides of this slightly artificial fence were very different. The study of the 'perception of the social' concerned itself mainly with the ways in which



people perceive other people; 'social factors in perception' came to mean the effects of a variety of socially derived influences on the perception of the physical environment. As Hochberg (1957) pointed out, the preoccupation of the psychologists studying these influences was with perceptual processes, and the principal object of their research to help elucidate some general problems in perception. The interests of those concerned with 'perception of the social' were in the fields of social behaviour and of personality, and their findings were used as a tool in the analysis of social processes, or for diagnostic purposes in clinical practice.

It must be added that this enticing symmetry is to some extent misleading. The rough correspondence of one aspect of work in social perception with perceptual interests, and of the other with social and personality interests, is by no means without exceptions. 'Perception of the social', or more simply, of people, was also seen by some as providing a testing ground for general perceptual theory, and in some cases, interests in perception and in personality have been fused in investigations concerned with the relationships between the two.

In a field as wide and varied as this, any choice of more or less representative experiments must be considered arbitrary and expendable: any number of other choices is possible and equally well justified. The experiments to be described below have been chosen from three areas of research indicated earlier:

- (i) The effects of social factors on perception of the physical environment;
- (ii) The effects of labels ('stereotyping') on perception of people;
- (iii) The nature of inference in perception of people.

SOCIAL FACTORS IN PERCEPTION

Research on the role played by social factors in the perception of physical environment concerned itself mainly with two problems: is perception of the physical attributes of an object affected by the value or the emotional significance to us of this object? And: how

is our perception of the physical world affected by the information that we receive about it from various social sources?

Emotional significance and value. The results of the first group of studies have been a subject of controversy for a number of years; this controversy cannot be gone into, or even less, resolved, in these pages. It seems, however, that two statements which have sufficient empirical support can now be made. First, under some conditions, value or the emotional relevance of an object may affect the perception or judgement of this object. The size of valued objects is sometimes over-estimated, or differences in value between a series of objects may induce an exaggeration of judged differences in size (or in other physical properties) between them (McCurdy, 1956; Tajfel, 1957). Secondly, it has been shown (see for example Singer, 1956, or Dixon, 1958) that the emotional significance of a stimulus may have an effect on its recognition and on some attendant phenomena: sometimes it increases the sensitivity, sometimes it prevents swift and efficient recognition.

This must not be taken to mean that the main effect of social factors is to produce 'distortions' in perception. Attending, or even sometimes 'refusing' to attend, to some selected aspects of the physical world may be useful, and can be conceived as one facet of the general phenomenon of perceptual selectivity. One of the prime conditions of survival is the capacity to deal selectively with the infinite amount of sensory information impinging continuously on the organism, and to 'notice' only those aspects of the world which are relevant at the time. Similarly, some shifts in the perception or judgement of size or of other physical properties of objects can be shown to serve a function: sometimes awareness of differences (and accentuation of these differences) between objects may turn out to be more important than minutely accurate judgements of their size.

Social and cultural consensus. What we *know* about an object, as distinct from what we see, hear, smell, or taste of it, affects the manner in which we perceive it. One of the important and reliable sources of knowledge about the physical world is the information about it transmitted through various social channels: consensus of opinion, tradition, language, etc.

The effects of social consensus on judgements of length have been investigated in the well-known studies by Asch (1952), followed by many others. These studies, conducted in the miniature context of small groups, have shown that many people are in a state of conflict when the information provided by their senses contradicts flatly the information that they receive from others. The situation was very simple: the majority of a group, in conspiracy with the experimenter, reported judgements of length which were obviously wrong. Judgements of length obtained from the unknowing 'victims' in the group provided the data. Information received from other people about the simple physical properties of objects is usually reliable and fits in with the sensory information. When the two flagrantly conflict, the least that happens is doubt, hesitation, and conflict; the most – a change in judgement in the direction of the group consensus.

Asch's experiments were conducted in an American college, and some tend to suspect these institutions of providing a favourable ground for development of strong tendencies towards conformity. However this may be, similar results have been obtained elsewhere. An unpublished experiment by Bullimore and the author conducted in an English school gave results closely agreeing with those reported by Asch. Milgram* has recently worked on the same problem (using estimates of duration of continuous tones) in Norway and in France, and found social pressure to be effective in both countries, though less so in France than in Norway.

Consensus of opinion in groups, small or large, is by no means the only social influence which comes to pattern our view of the physical world. Cultural differences in familiarity with objects and shapes are also reflected in some perceptual phenomena. At the turn of the century, Rivers (1904) has shown this to be the case in the perception of some geometrical illusions. Similar findings about an illusion the extent of which is thought to depend upon familiarity with windows and all sorts of other rectangular objects have been

* Milgram's work has not yet been written up in its final form for publication. The account of it is based on a letter from Milgram, summarizing his main results.

reported recently by Allport and Pettigrew (1957) in a study conducted in South Africa.

In summary, it can be said that when the experimental conditions are so arranged that a discrepancy exists between the two main sources of information about the physical world, the sensory and the social, this discrepancy reflects itself in a conflict. A prediction about the mode of resolution of this conflict - towards the social or towards the sensory - which would be valid in all circumstances, cannot be made. The perceptual end-product depends upon a number of attendant conditions, the most important of these being probably the clarity which with the sensory information is received.

Language. Languages differ not only in their grammar, syntax and phonetics, but also in their idiosyncratic ways of slicing the external world. Rivers (1904) found already that the Todas have '... a tendency to discriminate greens, blues, and violets less definitely than reds and yellows, and that the deficiencies in nomenclature for the former group of colours are accompanied by a certain degree of deficiency in their discrimination' (p. 328). He thought, however, that this was adding to the evidence '... in favour of the view that one of the factors upon which the defective colour nomenclature of the lower races depends is some degree of defect in sensitiveness' (p. 392).

Different languages tend to impose different groupings on such physical 'continua' as the range of colours. These groupings can be described in terms of three principal characteristics: the number of labels assigned by a language to a continuum; the regions on the continuum where the assignment of one label tends to shade over into the assignment of another; and the consistency with which a particular label is assigned to a particular stimulus on the continuum. Lenneberg and Roberts (1956) provide an example of the first in comparing the number of temperature terms used in Portuguese and in English. Three terms exist in English: hot, warm, and cold. In Portuguese, there are only two: *quente* and *frio*, with the intermediate terms (*tepid* and *caldo*) not frequently used. Similarly, the French term *tiède* is more rarely encountered than *chaud* and *froid*. Differences in the number of terms used imply differences as to the regions

of the continuum to which these terms are applied. For example, Lenneberg and Roberts found that the Zuni Indians made one category of the colour region divided by the English speakers into yellow and orange.

The third characteristic is the consistency with which a name is applied to a stimulus. For example, in English the principal colour names in common use are: red, orange, yellow, green, purple, pink and brown. Each of these can be conceived as applying to a region of colours, but not as applying to each colour within a region with equal degree of consistency. Sometimes this is shown by the fact that for some colours we feel the need of specifying further beyond just calling them 'blue', 'red', or 'yellow'; sometimes by lack of agreement between observers about the name to be applied. Each colour region, however, has a 'focus' which can be specified both by the definiteness (i.e. lack of additional qualifications) of the name applied to it, and by the inter- and intra-observer agreement in the naming of it.

An experiment on colour recognition

This third characteristic, referred to as 'codeability' by Brown and Lenneberg (1954), offers the possibility of ascertaining the effects of language on recognition of colours, without recourse to cross-cultural comparisons. It is, as they say, 'a matter of comparing the English codeability of one region of visual experience with another region, whereas the ethnolinguist has usually compared the codeability of one region of experience in several languages. If we explore the codeability variable in English, it seems likely that our discoveries will apply to the cultural differences' (p. 458).

Preliminary procedure: the determination of 'codeability'.

A valid measure of codeability must be found in order to determine the effects of it on recognition of colours. This can be done in the following way: the subjects are shown for about five minutes a white chart on which twenty-four colour chips are mounted. In the experiment by Brown and Lenneberg, these colour chips were selected from the well-known Munsell system of colour specification in such a way that they covered more or less evenly the 'colour space'.

SOCIAL PERCEPTION

They were mounted on the chart in random order. Immediately after the five-minute exposure of the chart, the colours are presented again one by one in random order, and the subjects requested to name each of them in turn. Care must be taken to ensure that all the subjects have the same language background, and that none are colour blind. It was found in the original experiment that agreement between observers was the most useful index of codeability. This finding was confirmed in a subsequent study by Lenneberg (1957).

TABLE 1.¹ The Munsell notation and scores for discriminability, codeability, and recognition for the 24 test colours

<i>Munsell notation</i>	<i>Discriminability</i>		<i>Codeability</i>		<i>Recognition (Group C, Table 1)</i>	
	<i>Score</i>	<i>Rank</i>	<i>Score</i>	<i>Rank</i>	<i>Score</i>	<i>Rank</i>
2.5R 7/8	38	2	18	9.5	.875	8
2.5R 5/10	27.5	6	7	18.5	.694	11
5R 4/14	23	10.5	19	7.5	1.020	5
7.5R 8/4	18	15	7	18.5	.236	18
2.5YR 6/14	38	2	29	1.5	1.499	2
5YR 3/4	24	9	26	3	.972	7
7.5YR 5/8	26	7.5	8	16	.736	9
2.5Y 7/10	12	19	3	24	.486	13
5Y 8/12	37	4	25	4	2.450	1
7.5Y 6/8	13	17	4	23	.250	17
3GY 7.5/11.2	23	10.5	14	12	1.222	4
7.5GY 3/4	9.5	23	14	12	0.000	23.5
2.5G 5/8	18.5	14	23	6	.986	6
7.5G 8/4	17.5	16	19	7.5	.167	19
5BG 3/6	4.5	24	12	15	.111	22
10BG 6/6	21	12	7	18.5	.458	14
8.5B 3/6.8	38	2	13	14	0.000	23.5
2.5PB 7/6	19	13	18	9.5	.436	16
5PB 4/10	10.5	21	29	1.5	.695	10
10PB 5/10	12	19	7	18.5	.125	20.5
5P 8/4	12	19	14	12	.547	12
10P 3/10	10	22	24	5	.444	15
5RP 6/10	26	7.5	6	21.5	.125	20.5
8RP 3.4/12.1	31	5	6	21.5	1.464	3

¹ Reproduced from Brown & Lenneberg, 1954; Table 1, p. 459.

With the use of this criterion, the twenty-four colours can be ranked in order of their codeability. If Munsell colours are employed for repeating the experiment, there is no need to repeat this procedure, as the original ranking is available and can be used directly for the study of colour recognition. Columns 1, 4 and 5 in Table 1 set out the Munsell notation and the codeability scores and ranks for each of the twenty-four colours. If Munsell material cannot be used, the determination of codeability can be made by selecting twenty-four commercially available colours which represent more or less evenly the entire range of colours, and by repeating the Brown and Lenneberg procedure on a group of adult subjects who will not take part in the subsequent recognition experiment. The only precautions needed are again those of excluding subjects who are colour blind or who have different language backgrounds.

Main experimental procedure: correlating 'codeability' and recognition. For the study of the relationship between codeability and recognition, Brown and Lenneberg selected alternate chips from the 240-chip Munsell series at highest saturation, taking care to include the twenty-four chips used previously in the measurement of codeability. The resulting collection of 120 chips was mounted on a white chart. Table 2 sets out the disposition of the colours on the chart. Again, if Munsell colours are not available, the stimuli are easily manufactured and can be mounted on a chart in some predetermined order.

The task of the subjects in the main experiment is to recognize on the large chart colours which have just been shown to them; these are the twenty-four colours which have been previously ranked for codeability. Exposure time of the colours presented for subsequent recognition is three seconds, their order of presentation randomized, and the subjects are tested individually.

"The basic procedure was to expose simultaneously four of the twenty-four colours, remove them and ask the subjects to point to the colours just seen, on the large chart of 120. Neither the experimenter nor the subjects mentioned any colour name during the session" (p. 460).

TABLE 2.¹ Munsell notation of the disposition of colours on the Brown and Lennenberg chart

2.5G/8	5G/7	5G/6	2.5G/5	5G/4	2.5G/3	7.5G/3	10G/4	7.5G/5	10G/6	10G/7	7.5G/8
7.5GY/8	10GY/7	7.5GY/6	10GY/5	7.5GY/4	7.5GY/3	2.5BG/3	7.5BG/4	2.5BG/5	5BG/6	5BG/7	2.5BG/8
3GY/7.5	5GY/7	2.5GY/6	5GY/5	2.5GY/4	5GY/3	5BG/3	2.5B/4	2.5BG/4	10BG/6	10BG/6	7.5BG/8
10Y/8	7.5Y/7	7.5Y/6	10Y/5	7.5Y/4	10Y/3	10BG/3	7.5B/4	2.5B/5	5B/6	5B/7	2.5B/8
5Y/8	2.5Y/7	5Y/6	2.5Y/5	2.5Y/4	5Y/3	5B/3	2.5PB/4	7.5B/5	10B/6	10B/7	7.5B/8
7.5YR/8	10YR/7	10YR/6	7.5YR/5	5YR/4	10YR/3	8.5B/3	5PR/4	2.5PR/5	5PR/6	2.5PB/7	5PB/8
2.5YR/8	5YR/7	7.5YR/6	2.5YR/5	2.5YR/4	5YR/3	7.5PB/3	10PB/4	7.5PB/5	2.5P/6	7.5PB/7	10PB/8
7.5R/8	10R/7	2.5YR/6	7.5R/5	7.5R/4	10R/3	2.5P/3	5P/4	10PB/5	7.5P/6	2.5P/7	5P/8
5R/8	2.5R/7	5R/6	2.5R/5	5R/4	2.5R/3	7.5P/3	2.5RP/4.7	5P/5	10P/5	7.5P/7	10P/8
2.5RP/8	7.5RP/7	10RP/6	7.5RP/5	10RP/4	8RP/3.4	10P/3	5RP/3	2.5RP/4	5RP/6	2.5RP/7	5RP/8

¹ By courtesy of Dr. E. H. Lennenberg. Symbols are in terms of hue and brightness. Saturation to be used is the highest available.

"The recognition score for a colour was computed as follows: we determined the number of correct identifications made by each subject and considered this number to be unity. Each individual correct identification was given the appropriate fractional value" (p. 460).

Thus, the score for each colour correctly identified by a subject is expressed as a fraction of one having for its denominator the total number of colours correctly identified by this subject. In this way, the total recognition score for a colour is 'the sum of the individual performances weighted for each subject's over-all ability to recognize colours'.

A ranking order can thus be established for the ease of recognition of the twenty-four colours whose degree of codeability had been previously assessed. A significantly positive correlation between the two ranking orders would mean that codeability is a determinant of the ease with which a colour is recognized.

Brown and Lenneberg were interested in the testing of one further hypothesis: that the effect of codeability on recognition would become more marked as the task of recognition was made more difficult. For this reason they conducted their experiment on four groups of subjects. The difficulty of the task was varied from group to group by varying the number of colours simultaneously exposed, the length of the intervals between exposure of the colours and the request to identify them on the chart, and by asking the subjects in one of the groups to engage in some irrelevant tasks during the intervals. The experimental conditions for the various groups are summarized in Table 3.

TABLE 3.¹ Recognition procedures

Group	Number of colours originally exposed	Length of interval	Content of interval
A	1	7 seconds	
B	4	7 seconds	
C	4	30 seconds	
D	4	3 minutes	Tasks

¹ Reproduced from Brown & Lenneberg, 1954; Table 3, p. 460.

A demonstration of the main effect can, however, be made by using one only of the conditions ascribed to the original groups B, C or D.

The problem of discriminability. The interpretation of the results of the experiment presents some difficulties. Even if the obtained correlations between codeability and recognition turn out to be statistically significant, they do not justify the conclusion that the two are in fact related to each other. There is another factor in the situation which might conceivably influence the results, and must therefore be taken into account: the colours displayed on the large chart may differ from one another in the ease with which they can be discriminated from the surrounding colours (their 'discriminability'). The efficiency of the recognition of a colour may thus depend not only on its codeability but also on its salience on the chart. If for example it is found that discriminability is related both to codeability and to ease of recognition, then a correlation between the latter two may not be due to a 'genuine' relationship between them, but may be simply expressing the fact that they have a third underlying common property.

A direct experimental determination of colour discrimination is beset with technical difficulties. Fortunately, Brown and Lenneberg were able to make use of some already existing assessments of discriminability applying to Munsell charts (Newhall, Nickerson, and Judd as quoted by Brown and Lenneberg). They ranked the twenty-four colours in order of their discriminability, and correlated this ranking order with their own measures of codeability. The resulting correlation (0.074) was too small to affect in any important way the relationship between codeability and recognition. Discriminability scores and ranks are set out in columns 2 and 3 of Table 1, and can be used directly for repetitions of the experiment with Munsell charts. If other colours are used, these discrimination measures cannot, of course, be applied. However, since the correlation between codeability and discriminability obtained by Brown and Lenneberg was minute, the chances are that it would not be very large with another set of colours, and that therefore it would not affect the results in any important way.

Results. Table 4 reproduces the main results obtained by Brown and Lenneberg.

TABLE 4.¹ Correlations involving scores on codeability (C), discriminability (D), and recognition (R) with four experimental conditions for recognition

Group	<i>C with R</i>	<i>D with R</i>	<i>C with R, D constant</i>
A	.248	.540*	.248
B	.411	.460*	.426*
C	.415	.503*	.438*
D	.487*	.505*	.523*

* $p \leq .05$

¹ Reproduced from Brown & Lenneberg, 1954; Table 4, p. 461.

Column 4 shows the partial correlations between codeability and recognition with discriminability held constant. As can be seen, the hypothesis that recognition becomes easier as a colour becomes more 'codeable' is in general confirmed. The only exception is for group A for which, it will be remembered, the conditions for recognition were the easiest. This exception and the trend of results for the other groups confirm the subsidiary hypothesis that the facilitating effect of codeability on recognition increases as the task of recognition is made more difficult.

LABELS AND PEOPLE

The number of ways in which people can be classified is infinite. Every individual has his own idiosyncratic ways of indulging in this game, and these depend on his past experience, interests, needs, and purposes. A classification may be permanent or transitory; it may be rooted in deep emotional involvement, or it may be emotionally neutral; it may be prominent, in the sense of being used as one of the first things to think about when meeting people, or used only on some very specific occasions. All these factors have their importance in an assessment of the consequences that classifications have in the perception and evaluation of people.

The essential questions to be asked about these consequences are:

what are the attributes that are inferred from the assignment to a class? How much are we prepared to say about an individual *because* he is a professional boxer, an accountant, a member of an anti-vivisectionist society, or a West Indian? To what extent can our ideas about him, or about the class of which he is a member, be changed by evidence that contradicts them? How have we come to associate certain attributes with a certain class of people?

The so-called 'national stereotypes' offer the classical example of inferring the characteristics of an individual from his membership of a group. They are, however, by no means the only one. Our knowledge that someone belongs to a professional, religious or social group often leads to inferences of a similar kind. Advertisers are well aware of the importance of a 'brand image': the creation of a public response to a name of a product, or a series of products, which consists of associating certain attributes with the name. Brand images are the stereotypes of commercial products; the stereotypes are the brand images of human groups.

The manner in which stereotypes function can be conceived to depend upon at least four sets of conditions.

(1) *The balance between the amount of specific information about an individual and of general information about his class:* as in the case of more general perceptual and judgement phenomena, we tend to use whatever sources of information happen to be available about the outside world. In the relative absence of clear-cut knowledge about someone, we tend to fall back on what we know about 'his kind', and to exaggerate the definiteness and the certainty of whatever we do know. Because of this, studies such as Katz and Braly's (1933) in which subjects were asked to characterize various national groups by 'checking' some adjectives in a long list, are probably not very good at predicting attitudes towards concrete individual members of the groups so characterized. These studies are, in a sense, experiments in a vacuum, and the responses cannot be assumed to stand in direct relation to those that would have been made if the subjects had a modicum of specific information about an individual. They have, however, a function: they elicit the 'cultural' stereotypes about national groups, even those entirely unknown to the subjects ('Turks

are cruel'), embedded in language, tradition, literature and history; and they do provide some hints.

(2) *Past experience*: the generalizations we form about a human group may be due to some consistencies, observed in the past, in the behaviour of various members of this group. These generalizations may be valid to some extent, but very often they tend to be applied in excess of their validity. Further, they need not rely on personally observed consistencies in the behaviour of members of a group. Newcomers and children are faced with the already existing generalizations, accept them and act on their basis. Knowledge about the social world surrounding us is partly based on social consensus, just as is our perception of the physical world; but, as it was already pointed out, social consensus increases in importance as other kinds of information become ambiguous and inconsistent. It will therefore play a predominant role in the perception and evaluation of people, as most characteristics of individuals acquire content and meaning only in relation to the social world of which they are a part.

(3) *Conformity*: social consensus in the form of influences exercised by the family, the school, the mass media of communication, the denominational, professional and other groups, the social, cultural and linguistic traditions, surrounds the individual. Pressures towards conformity serve to increase the impact of all these influences. It is altogether easier to say and to 'see' (as Asch has shown) what others say and see than to be different. A newcomer to the American South tends to acquire the Southern stereotypes about the Negro. The same may be true of a newcomer to Notting Hill.

This does not, however, account fully for the sharp edges that many stereotypes show, and for the infallible speed with which they are adopted in situations of violent tensions between groups. A distinction must be made between 'lazy' and detached conformity in mildly antagonistic situations, and conformity in situations marked by hatred, violence and riots.

(4) *Emotional involvement*: the extreme character of some stereotypes, the readiness with which they are used, the difficulty of changing them in face of contradictory evidence, cannot be entirely understood on the basis of their origin in past experience and in conformity.

Perception of the social world is, more even than perception of the physical world, affected by motivation. Selecting, emphasizing, omitting, accentuating differences are phenomena common to the orientation towards both. Thus, an intense emotional involvement in prejudice is accompanied by strong and unshakeable stereotypes.

Though differences in the degree of emotional involvement may account for individual differences in the strength and nature of stereotypes, some other factors must be responsible for the recurrent emergence, at various times and in various places, of attitudes towards a selected group which are shared by a mass of people, and which fit almost perfectly some clinical descriptions of paranoia. It can hardly be assumed that a majority of people living in a country, a district, or a city have suddenly acquired a personality structure described by Adorno *et al.* (1950) as 'authoritarian', or that a majority of them were latently authoritarian, waiting only for an occasion to demonstrate these latent tendencies. It seems more likely that tense situations require highly stringent conformity, and that the content of stereotypes changes in such cases simultaneously with the strength of social pressures to accept them. This is an explosive mixture, where the reliance on specific information about an individual is drastically reduced, and the label takes precedence on almost all occasions.

An analysis of stereotypes must start with the statement that an attribute assigned to an individual on the basis of his membership of a group does not float unanchored in a universe of absolute assertions. A man is not 'tall' in an absolute sense; he is *taller* than other people. Madame Butterfly (and, with her, other Japanese women) seems 'dainty' because the usual run of conventional heroines is definitely less dainty. Attributes assigned to people are arranged on a continuum from more to less, and have very little meaning outside their comparative framework. Therefore, the primary consideration in the study of stereotypes must be an assessment of the perceived or judged *differences* between groups of people. These differences can be predicted if we know the ways in which an individual tends to classify people, and the importance to him of the various classifications that he uses (Tajfel, 1959a).

These differences can only be studied and assessed in terms of departures from some known criteria. In the case of the effects of group pressures on judgement of length, the criterion is easy to find; there is no difficulty in measuring length. 'Honesty', or 'pleasantness', or 'cruelty' of a national, religious, or socio-economic group are not so easily dealt with. There are, however, two general ways of overcoming this difficulty: one consists of eliciting the responses of different groups of subjects to the same stimuli; the other, conversely, of keeping a group of subjects constant, while varying the name, label, or class assigned to a stimulus.

In the first category of studies, the main problem is to find out whether different groups of subjects tend to differ consistently in their assessments of the same people. Secord, Bevan, and Katz (1956) divided their subjects into groups, according to previously ascertained extent of anti-Negro prejudice, and asked them to judge the personality characteristics and the degree of physical negroidness of an assortment of photographs of 'white' and Negro faces. They found that, in their ratings, 'the anti-Negro judges exaggerate the personality stereotype of Negroes, whereas pro-Negro judges de-emphasize it' (p. 82). The prejudiced subjects also tended to perceive the Negro photographs as physically more negroid than the 'neutral' subjects did. Thus, both in their ratings of personality traits and of physical characteristics, the prejudiced subjects tended to accentuate the differences between the two classes of people more than the non-prejudiced subjects.

Applied in a different way, the procedure of having different groups of subjects characterize the same social group or groups yielded some interesting findings about historical changes in the 'cultural' stereotypes, discussed previously. The Katz and Braly study on ethnic stereotypes, conducted in 1933 on Princeton students, was repeated by Gilbert (1951) eighteen years later on a similar group of students. Gilbert found that, in general, the stereotypes were similar to those described in the previous study, but that, as might have been expected, the Japanese and the Germans appeared in a new guise to the post-war generation; but he also found that there was altogether less readiness to express generalizations about

national groups. A number of other studies have also shown modifications in stereotypes occurring as a consequence of changes in political, economic, and social conditions.

The studies just described provide information about differences in the attitudes of different groups of people towards a specific group. This is complementary to assessing the differences in the attitudes of the same people towards different groups. However, using for this purpose 'genuinely' different people as objects of judgement presents an obvious difficulty: the differences in judgment may be due not to the subjects' tendencies to assess an individual according to his membership of a class of people, but to some real differences between the objects of judgement. This can be overcome by presenting the same individual, characterized at various times as belonging to one or another class. This device has been used not only by psychologists. The hero in Arthur Miller's novel *Focus* (1945) is a very short-sighted individual who, when he wears glasses, looks like a Jew (for purposes of the plot he cannot bear wearing contact lenses). He finally must resign himself to wearing glasses continuously, and bear the consequences. The action takes place in the thirties, when anti-semitism tended sporadically to become an acute problem in some localities in America. The change in the man's appearance causes his employers and neighbours to believe that he was a 'hidden' Jew who now stands revealed, and the book provides an excellent study of the sudden change in their attitudes towards him, and in their interpretation of practically everything he says or does.

Razran's (1950) study of ethnic stereotypes is an experimental and less dramatic counterpart of Miller's situation. He asked his subjects to rate a number of photographs on several personality traits. Some time later, the same subjects were asked to rate the same photographs (to which others were added in order to prevent recognition), but this time each photograph had an ethnic label. A comparison of the two ratings showed a shift in judgement in the direction determined by the stereotypes of the groups to which the individuals on the photographs were now assumed to belong.

An experiment on the effects of stereotypes on judgement of personality

Lambert, Hodgson, Gardner, and Fillenbaum (1960) combined in their experiment the two procedures described above. The investigation was conducted in Montreal, and their aim was to elicit the English and French stereotypes of both the English and French sections of the community. As will be seen, the procedure of the Canadian experiment can be applied to a variety of other situations in which stereotypes are functioning.

Procedure. The subjects of the original experiment were groups of English- and French-speaking Montreal students, of similar age and educational level. They were asked to evaluate the personality characteristics of four bilingual speakers who recorded on tape French and English versions of a two and a half minute passage of prose. The subjects were not aware that each speaker read the passage in both languages, '... so that the evaluational reactions to the two language guises could be matched for each speaker'. The passage, originally French, was 'of a philosophical nature'. The recordings, preceded by two practice trials of other speakers' recordings, were presented 'in alternating French-English order, allowing the maximum possible interval between successive presentation of the English and French guises of any speaker'. The study was presented to the subjects as being concerned with the extent to which people's judgements about an individual are determined by his voice. The subjects were given previously prepared response sheets, and their task was to rate each of the voices on fourteen personality traits, each trait on a six-point scale ranging from 'very little' to 'very much'. Each voice was played once, and the rating of it was done 'both while it was playing, and during the ninety second interval between voices'. At the bottom of each response sheet, the subjects were asked to indicate what was the likely occupation of the owner of the voice.

The four speakers who recorded the passage spoke both languages in faultless idioms, and had in the past used both in habitual, everyday situations. All four spoke English with the standard Canadian

accent, two ('Bla' and 'Cou') spoke French with the French-Canadian accent, one ('Leo') 'with a marked French-Canadian accent characteristic of those who work in the "bush"', and one ('Tri') with a French accent from France.

'For each subject on each of the fourteen traits the difference (D) between the evaluations of the English and French guises of each speaker was noted, with account taken of the direction of the difference. These D values were then summed over the four speakers and over the subjects.'

Results. The over-all results of the ratings, showing for each group of subjects the differences between their evaluation of the French and English guises of the same speakers, are summarized in Table 5. As can be seen, the English group finds the English voices superior to the French on seven traits out of fourteen: height, good looks, intelligence, dependability, kindness, ambition, and character; the French superior to the English in sense of humour only. The French group finds the English superior to the French on ten traits out of fourteen: height, good looks, leadership, intelligence, self-confidence, dependability, ambition, sociability, character, and likeability; the French superior to the English in religiousness and kindness.

Adaptation of the experiment to other situations. Lambert's method offers an excellent technique for dragging into the open some more or less implicit stereotypes. It combines the procedures of keeping the stimulus constant and varying the groups of subjects (comparisons between the English and French reactions to the same voices), and of changing the designation of the stimulus and keeping the group of subjects constant (comparisons within the French or English groups of reactions to the same voice speaking one or the other language). This could be repeated in the classroom by making tape-recordings of the same person's voice when he is using different accents or dialects.

Discussion of Lambert's results. In a number of ways Lambert's results seem odd: the English tendency to prefer the English guises of the speakers is less pronounced than the French tendency to do precisely the same thing. This does not conform to our stereotype

An experiment on the effects of stereotypes on judgement of personality

Lambert, Hodgson, Gardner, and Fillenbaum (1960) combined in their experiment the two procedures described above. The investigation was conducted in Montreal, and their aim was to elicit the English and French stereotypes of both the English and French sections of the community. As will be seen, the procedure of the Canadian experiment can be applied to a variety of other situations in which stereotypes are functioning.

Procedure. The subjects of the original experiment were groups of English- and French-speaking Montreal students, of similar age and educational level. They were asked to evaluate the personality characteristics of four bilingual speakers who recorded on tape French and English versions of a two and a half minute passage of prose. The subjects were not aware that each speaker read the passage in both languages, '... so that the evaluational reactions to the two language guises could be matched for each speaker'. The passage, originally French, was 'of a philosophical nature'. The recordings, preceded by two practice trials of other speakers' recordings, were presented 'in alternating French-English order, allowing the maximum possible interval between successive presentation of the English and French guises of any speaker'. The study was presented to the subjects as being concerned with the extent to which people's judgements about an individual are determined by his voice. The subjects were given previously prepared response sheets, and their task was to rate each of the voices on fourteen personality traits, each trait on a six-point scale ranging from 'very little' to 'very much'. Each voice was played once, and the rating of it was done 'both while it was playing, and during the ninety second interval between voices'. At the bottom of each response sheet, the subjects were asked to indicate what was the likely occupation of the owner of the voice.

The four speakers who recorded the passage spoke both languages in faultless idioms, and had in the past used both in habitual, everyday situations. All four spoke English with the standard Canadian

accent, two ('Bla' and 'Cou') spoke French with the French-Canadian accent, one ('Leo') 'with a marked French-Canadian accent characteristic of those who work in the "bush"', and one ('Tri') with a French accent from France.

'For each subject on each of the fourteen traits the difference (D) between the evaluations of the English and French guises of each speaker was noted, with account taken of the direction of the difference. These D values were then summed over the four speakers and over the subjects.'

Results. The over-all results of the ratings, showing for each group of subjects the differences between their evaluation of the French and English guises of the same speakers, are summarized in Table 5. As can be seen, the English group finds the English voices superior to the French on seven traits out of fourteen: height, good looks, intelligence, dependability, kindness, ambition, and character; the French superior to the English in sense of humour only. The French group finds the English superior to the French on ten traits out of fourteen: height, good looks, leadership, intelligence, self-confidence, dependability, ambition, sociability, character, and likeability; the French superior to the English in religiousness and kindness.

Adaptation of the experiment to other situations. Lambert's method offers an excellent technique for dragging into the open some more or less implicit stereotypes. It combines the procedures of keeping the stimulus constant and varying the groups of subjects (comparisons between the English and French reactions to the same voices), and of changing the designation of the stimulus and keeping the group of subjects constant (comparisons within the French or English groups of reactions to the same voice speaking one or the other language). This could be repeated in the classroom by making tape-recordings of the same person's voice when he is using different accents or dialects.

Discussion of Lambert's results. In a number of ways Lambert's results seem odd: the English tendency to prefer the English guises of the speakers is less pronounced than the French tendency to do precisely the same thing. This does not conform to our stereotype

TABLE 5.¹ *t* values for significance of differences in evaluations of English and French speakers.²

Trait	Speakers:	English subjects' judgements					French subjects' judgements				
		Cou	Bla	Leo	Tri	Over-all	Cou	Bla	Leo	Tri	Over-all
Height		8.63*	4.67*	4.05*	2.59†	8.83*	5.50*	2.32†	2.41†	1.04	4.58*
Good looks		7.16*	4.10*	2.66*	-.90	5.78*	7.50*	6.00*	7.26*	1.74	9.77*
Leadership		1.72	.69	1.41	-1.82	1.20	4.05*	9.32*	12.10*	.82	11.06*
Sense of humour		-.58	-.14	2.25†	-3.48*	-2.16†	2.67*	.58	.64	-.30	1.40
Intelligence		1.44	.11	2.25†	1.98†	2.40†	4.54*	9.17*	11.73*	1.00	10.22*
Religiousness		-.67	-.33	-.26	1.81	.27	-1.79	-3.11*	-1.60	-1.14	-2.94*
Self-confidence		-.50	-.47	1.13	-1.33	-.69	5.37*	7.56*	10.00*	-2.00†	8.28*
Dependability		1.61	-.15	3.51*	2.74*	3.11*	2.89*	4.36*	7.50*	1.89	6.44*
Entertainingness		1.77	1.67	.00	-4.44*	-.84	.44	-.27	-.81	.22	-.21
Kindness		2.40†	2.47†	1.67	.74	3.37*	.10	-4.12*	-2.65†	-.74	-2.96*
Ambition		.45	2.43†	2.33†	.84	2.83*	3.06*	4.64*	7.53*	-1.53	5.89*
Sociability		.24	-.16	-.17	-4.13*	-1.44	3.44*	2.10†	3.47*	.29	3.72*
Character		3.07*	1.00	3.17*	.12	3.08*	3.95*	6.81*	8.68*	.31	8.51*
Likeability		2.38†	.26	1.38	-1.21	1.22	1.83	.80	2.05†	.79	2.00†

¹ Reproduced from Lambert *et al.* 1960, Table 1.² Positive entries indicate that English guises are evaluated more favourably than French, and minus entries indicate more favourable evaluation for French guises. This is so for both French and English subjects' judgements.

† significant at 5 per cent. level, two-tailed tests. * significant at 1 per cent. level, two-tailed tests.

of stereotyping which usually carries the implication of favouring one's own group. Lambert *et al.* report that none of the questionnaires which were completed by their subjects yielded convincing relationships with the data from ratings of voices, and for an explanation of their results fall back on the idea of 'self-hate': the tendency, shown sometimes by the minority groups, to accept the majority's unfavourable characterization.

An alternative explanation is possible (Tajfel, 1959b), and has the advantage of accounting for the fact that the French subjects accentuate the differences in favour of the English voices *more* than the English subjects do in some traits, and not in others. Stereotypes were discussed above as expressing judgements of differences between human groups, especially of those differences which are of relevance to the observer. A separate inspection in Table 5 of preferences expressed by both groups of subjects for one or the other incarnation of each of the speakers shows that the French subjects tend to accentuate more than the English ones the English superiority in traits related to socio-economic success. The English section of the community does tend in Montreal to be socio-economically superior to the French one; the subjects knew it, as is shown by the manner in which they ascribed 'likely' occupations to the voices. It can be assumed that this situation is of direct and serious concern to the French subjects, especially as they were drawn from the college population. It is also interesting to note that in the case of the speaker 'Tri', the voice whose French accent was 'Parisian', the French subjects did not think that his English counterpart was better at anything than he was.

Two further points should be made about the Lambert study, both limiting its generality of application. There is no doubt that the authors have been able to establish the validity of their claim that 'evaluational reactions to a spoken language' are closely related to the general stereotype of the group that speaks that language. At the same time, it is not inconsistent with Lambert's findings to point out that the 'situational balance' will work in his experiment in the direction of making the stereotype more pronounced than it might have been in 'real life'. Voice was the only source of information

that the subjects had about the individuals whom they were assessing; normally, it is by no means the most important one. Therefore, we are confronted here once again with a situation in which paucity of specific information about an individual causes the subjects to fall back on general information about the class to which he belongs. This might have been further accentuated by the instructions to the subjects, informing them that the experiment was on people's ability to judge personalities from voices.

The second point is related to the behaviour of the speakers rather than of the subjects. We do not necessarily 'feel' the same way when we speak different languages or use different accents. This becomes quite clear when one listens to any good actor whose part requires the use of an accent which is not normally his. We do not know how Lambert's speakers felt about using French or English. If they shared the generally accepted stereotypes (which is likely), they might have 'adopted' to some extent the personality fitting with those stereotypes when speaking the language which was not originally theirs. This might have affected the extent of stereotyping in the subjects' ratings. It is difficult to know how to deal with this in experiments using language or accent as a tool for eliciting stereotypes. Possibly, it could be solved by gathering information about the speakers' own attitudes, and by using a large number of speakers with varied backgrounds.

INFERRING PERSONALITY TRAITS

Inferring personality characteristics of an individual on the basis of usually insufficient information can be considered 'perceptual' only by virtue of the fact that this activity has been often referred to as 'social perception', 'person perception', or 'perception of people'. There are, of course, some important similarities of process with perception proper, and they are mainly in the type, and even more in the vicissitudes, of inference occurring in both activities. Jumping to conclusions, or 'going beyond the information given,' is a common characteristic of most cognitive processes (Bruner, 1957), and it is extremely well exemplified in the private little 'per-

sonality theories' that we form about practically everyone within reach.

This is shifting ground for systematic enquiry. In a review of a recent symposium on the subject, Cohen (1959) writes that the book is '... a welcome sign that at least some psychologists are now turning to the study of questions recognized as psychologically significant in everyday life'; but, he adds, 'the volume as a whole can hardly be said to leave the reader with an impression that very much has been achieved' (p. 82). This seems a fair comment on the present state of this field, though, hopefully, not yet an epitaph. Some promising beginnings have been made; but a brief review must be content to deal with questions rather than with answers to them.

The problems seem to be of four types:

(1) *What is the basis of inference about other people?* The information that we have about others consists mainly of how they look and sound, and of what they do. Our conclusions about them must stand in some relation to this information; some characteristics must be selected as important, some omitted; some given greater weight than others for making predictions about future behaviour. Brunswik's (1956) work on the relationship between physical cues and the inferences drawn from them is relevant here. This type of investigation contains some promise of slow and painful progress, if it can be replicated again and again for large numbers of subjects and conditions. Its interest then may not be as much in discovering *what* cues are used for drawing inferences (with some insight each of us can discover this for himself), as in providing information about the misuse of these cues. As in the case of labels discussed in the previous section, the more ambiguous is the situation, the more we tend to rely excessively on whatever we happen to know, or think we know: certain slight relationships, such as between intelligence and the height of forehead, or perhaps between a booming voice and ability for leadership, assume in our judgement an importance entirely out of proportion with their degree of trustworthiness.

(2) *What are the relationships between an individual's personality, aims, and interests, and the inferences that he draws about others?* In the past, this question has been mainly of clinical interest. Here

again, the most valuable information has been gathered from glaring and consistent *misjudgements*. The Freudian and allied concepts of projection have been at the basis of much of the work in this area, and some experiments conducted with a moderate degree of success.

Some recent developments indicate, however, that the relationships between the individual and what he sees in other people may soon come to be viewed in a context wider than that of projection and of clinical diagnosis. Hastorf, Richardson, and Dornbusch (1958) have entered a plea for a closer consideration of the problem of 'relevance' in the study of person perception. They argue that many of the studies in which the subjects are requested to describe other people in terms of some attributes selected by the experimenter suffer from a lack of realism. Different people use different categories for descriptions of others – the main determinants of choice of categories being their relevance to the individual's aims, to the situation in which he finds himself, and also their continuity with the individual's past experience. One of the principal tasks of research should therefore be the analysis of the type and content of categories used by people when they confront various situations and various classes of other people. In brief, Hastorf *et al.* feel that there is a need for studying 'the qualities of a person's experience of others in terms of the verbal categories he uses in reporting that experience' (p. 56).

(3) *What are the situational determinants of inferences about others?* The importance of situations in inferences about others is already clear in Hastorf's considerations about 'relevance'. The term 'situation' is used here in a wide sense, including group or groups of which an individual is a member, his position in these groups, and his relationships with other group members. Much of the work on these situational determinants falls into the category, mentioned in the Introduction, of research in person perception primarily based on interest in social interaction. The findings have been applied with a fair degree of success, to concrete situations, with the aim of predicting on the basis of information about the individuals' perceptions of one another such group characteristics as efficient or inefficient

leadership, the stability of a group, or its fitness for the task which it is purported to fulfill.

(4) *How does the image of the 'whole' result from its component parts?* This is, of course, the perennial problem in the psychology of perception, raising its head in a new context. Asch (1946) was probably the first to apply to this aspect of it a thoroughgoing Gestalt solution. He maintained that our total impression of people is not the 'sum' of part impressions. His experimental technique consisted of asking his subjects to rate on a 'check-list', consisting of eighteen traits, an individual who, they were told, was 'intelligent, skilful, industrious, warm, determined, practical and cautious'. Another group of subjects rated on the same check-list another individual who was similar to the first in all respects, apart from being 'cold' instead of 'warm'. The two groups of subjects produced very different impressions of the two individuals. On the other hand, varying 'polite-blunt' instead of 'warm-cold' did not determine any important differences in the ratings. Asch's conclusions were that the final impression can be understood only as a consistent and unified image whose 'parts' acquire their meaning only in terms of their context, and that for the formation of this final impression some traits are more important than others ('warm-cold' *vs.* 'polite-blunt') because of their central rather than peripheral character in the total structure.

Wishner (1959) undertook a further analysis of this problem. Whether a trait is central or peripheral is not, according to him, unpredictable but depends on the manner in which traits tend to be related to each other when they are generally used. Wishner first established the intercorrelations between the various traits used in Asch's study by having 214 subjects rate a few individuals on all of them. The results showed that 'warm-cold' correlated most highly with those items on Asch's check-list which had shown the greatest differences between the 'warm' and the 'cold' groups. This result, Wishner writes, 'indicates that one can predict results in the Asch-type situation from a *prior*, independent knowledge of the relationships obtaining between the traits to be rated and the stimulus traits'. He further substantiated his conclusions by varying 'intelligent-unintelligent', and obtaining again differences in ratings predictable

from previously known correlations between the new pair and the traits on the check-list; and by showing that 'blunt-polite' or 'humane-ruthless' could be made either more 'central' or more 'peripheral', depending on whether their correlation with the traits selected for the check-list were high or low.

An experiment on inference of traits

Using different procedures, Bruner, Shapiro, and Tagiuri (1958) obtained results generally consistent with those reported by Wishner. The question they asked was almost identical: 'Can we predict from knowledge of inferences made from single traits the inferences that will be made when these traits are combined?' (p. 278). The object of their study was 'to inquire into the relation between inferences made from single traits in *isolation* and the inferences made from the same traits in *combination* . . . If we know a man is *considerate*, we infer he is *kind*; not so an *irresponsible* man. But when a man is both *considerate* and *irresponsible*, is he kind or is he not?' (pp. 278-9).

Procedure. Two groups of traits are used in the study: the 'given traits', from which inferences to others have to be made; and the traits which have to be inferred from the 'given' ones. There are four given traits: *intelligent*, *considerate*, *independent*, *inconsiderate*, and all combinations of these (excluding those employing *considerate* and *inconsiderate* simultaneously): *intelligent and considerate*; *intelligent and independent*; *intelligent and inconsiderate*; *considerate and independent*; *independent and inconsiderate*; *intelligent, independent, and considerate*; *intelligent, independent, and inconsiderate*.

Each subject is given a list of 59 traits (see Table 6) and asked to indicate to what extent people who have *one* of the single given traits or of their combinations, would also possess any of the 59 traits on the list. Thus, a subject might have a list at the head of which appeared the words: 'People who are *considerate* . . . very often are . . . tend to be . . . may or may not be . . . tend not to be . . . very seldom are . . . ' aggressive, awkward, active, etc. Each subject is working with one only of the possible eleven combinations of the given traits. In the original experiment each of the eleven combina-

tions was administered to 120 subjects (60 men and 60 women), and orders of the 'inferred' traits on the lists were randomized.

The Bruner *et al.* experiment is easily repeatable in the form of a small group project. Each member of the working group can gather data from as many subjects as possible, using one of the eleven combinations and different random orders of 'inferred traits' on his lists. Completion of the response sheet by a subject takes no more than ten to fifteen minutes. In this way, in a few days an array of data can be gathered, certainly not as large as in the original study, but sufficient for purposes of breakdown and analysis. The original data show very high agreement between the results obtained from men and women: it is therefore probably unnecessary to analyse them separately.

Results. As will be seen from Table 6, results are scored in terms of the amount of subjects' agreement in the inferences from given traits and from their combinations. Responses 'very often are' and 'tend to be' are considered as 'positive definite inferences'; those of 'tend not to be' and 'very seldom are' as 'negative definite inferences'; and those of 'may or may not be' as 'no inference'. The relations between the inferences from single traits and from their combinations were considered by Bruner *et al.* from three points of view: direction of inference; definiteness of inference; and the effects of consistency of information on the agreement between the subjects.

(a) Direction of inference: can inferences from a combination of traits be predicted on the basis of inferences from the same traits presented singly? If agreement is found amongst the subjects that someone who is inconsiderate is aggressive, and that someone who is intelligent is also aggressive, will they tend to think that someone who is inconsiderate and intelligent is aggressive? Examples of measure of agreement in such cases can be found in Table 6. The total inference about a trait was considered 'positive' when the majority of subjects used with regard to it the 'positive definite inference'; 'negative' when the converse was true. The general results concerning the direction of inference are presented in Table 7. As can be seen, the degree of predictability from inferences from single

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

TABLE 6.¹ Specific inferences made from *intelligent* (I) and *inconsiderate* (C) given singly and in combination (IC)

List traits	Given traits			List traits	Given traits		
	I	C	IC		I	C	IC
Aggressive	(+)	+	+	Imaginative	+	(-)	+
Active	+	(+)	+	Impulsive	(-)	+	+
Awkward	(-)	(+)	(-)	Indolent	-	(+)	(-)
Apathetic	-	(+)	-	Irritable	(+)	+	+
Boastful	(-)	+	+	Intellectual	+	(+)	+
Clever	+	(+)	+	Impractical	(-)	(+)	(-)
Considerate	(+)	-	-	Independent	+	+	+
Cold	(-)	+	+	Inefficient	-	(+)	(-)
Courteous	(+)	-	-	Inconsiderate	(-)	+	+
Unreliable	-	+	(+)	Loquacious	(-)	(+)	(+)
Unsympathetic	(-)	+	+	Neat	(+)	(-)	(+)
Cheerful	(+)	(-)	(-)	Non-intellectual	-	(-)	-
Deliberate	+	(+)	+	Practical	(+)	(+)	(+)
Dishonest	-	(+)	(-)	Modest	(+)	-	-
Dependent	-	(-)	-	Placid	(-)	-	-
Disorderly	(-)	(+)	(+)	Unintelligent	-	(-)	-
Discourteous	(-)	+	+	Reserved	(+)	(-)	(-)
Enterprising	+	(+)	+	Reliable	+	-	(-)
Even-tempered	(+)	-	-	Sympathetic	(+)	-	-
Efficient	+	(-)	+	Responsible	+	-	(-)
Energetic	+	(+)	+	Sociable	(+)	(-)	(-)
Enthusiastic	(+)	(-)	(+)	Sincere	(+)	-	(-)
Conscientious	+	(-)	(+)	Submissive	-	-	-
Friendly	(+)	(-)	(-)	Sad	(-)	(-)	(-)
Honest	+	(-)	(+)	Taciturn	(-)	(-)	(-)
Hostile	(-)	(+)	(+)	Unenterprising	-	(-)	-
Intelligent	+	(+)	+	Warm	(+)	-	-
Irresponsible	-	+	(-)	Witty	(+)	(+)	(+)
Hypocritical	(-)	+	(+)	Unimaginative	-	(+)	-
Humourless	(-)	(+)	(-)				

¹ Reproduced from Bruner, Shapiro, and Tagiuri, 1958, Table 1, p. 281.

NOTE: The entries in each column are based on an independent group of 120 subjects.

The symbols used in the table indicate the following:

- +
- (+) 50 per cent. or more of the subjects made positive definite inferences.
- (+) Less than 50 per cent. of the subjects made positive definite inferences but more of the definite inferences were positive than negative.
- 50 per cent. or more of the subjects made negative definite inferences.
- (-) Less than 50 per cent. of the subjects made negative definite inferences but more of the definite inferences were negative than positive.

TABLE 7.¹ Sign of inferences made from single traits and from their combinations

Sign of inferences made from single traits	Number of given traits	Sign of combination						Over-all total
		Predictable			Unpredictable			
		Men	Women	Total	Men	Women	Total	
Like (++ or --)	2	193*	194	387	2	1	3	390
Like (+++ or ---)	3	66	64	130	0	1	1	131
Unlike (+- or -+)	2	0	0	0	100	100	200†	200
Unlike (++ -- or -- +)	3	41	42	83	11	11	22	105
Total		300	300	600	113	113	226	826

¹ Reproduced from Bruner, Shapiro, and Tagiuri, 1958, Table 2, p. 282.

* Entries in the table stand for instances of trait combination.

† Of these 200 inferences, 90 were positive, 110 negative.



traits to inferences from the same traits combined in twos or threes is impressive: when inferences from single traits are of the same sign ('positive definite inference' or 'negative definite inference'), inferences from their combinations are predictable almost without exception.

In addition, efficient prediction can be made even in cases where inferences from single traits are in conflicting directions. When three traits which singly do not all yield inferences in the same direction are combined, the sign of the inferences from their combination is the same as the sign of the predominating inferences from single traits (i.e. two out of three) in 79 per cent. of cases (see Table 7). When two traits which singly lead to conflicting inferences are combined, the result of this combination can still be predicted. The basis for prediction is the amount of agreement achieved by the subjects for each of the single traits: 'agreement is here defined as the number of subjects making inferences in the direction of the dominant sign' (p. 282). Bruner *et al.* report that in 195 out of 200 such cases, the inference from a combination of two traits was of the same sign as the inference from the trait with the higher degree of agreement. They add that out of 22 instances of three-trait combinations (see Table 7) 'where the one single inference of sign unlike the two others determined the sign of the combination . . . in all but three instances the single inference was markedly higher in agreement than the remaining two' (p. 283).

In summary, the direction of inference from combined traits was predictable from the direction of inference in isolated traits in 97 per cent. of cases. Most of the remaining 3 per cent. were also predictable with the help of the additional criterion of agreement.

(b) In considering definiteness of inference, Bruner *et al.* were not concerned with the consensus in the sign of inference. Their aim was to find out how a combination of 'given' traits compares with a single 'given' trait in determining the subjects' readiness to make a definite statement about an inferred trait; in other words, to compare them with regard to the frequency of responses outside the null category of 'may or may not be'. They report that in 91 per cent. of cases when the traits singly 'lead to inferences of like sign', the

inferences from their combinations were more definite – i.e. showed a larger number of ‘positive’ or ‘negative’ statements than the mean number of such statements for the same traits in isolation.

(c) A final question asked by Bruner *et al.* was whether the *amount* of agreement is greater for inference from combined given traits of like sign than for inference from the components of the combination. Here again, they found that in 90 per cent. of such cases there was ‘an increase in agreement over the mean agreement of the component single inference’ (p. 285).

The massive support found in the data for the hypothesis of ‘predictability’ allowed Bruner *et al.* to bypass the statistical analysis of significance, and to rest their case with the simple counting of heads. If a repetition of the experiment does not yield results as clear-cut as those of the original study, it is, of course, always possible to resort to the usual techniques of frequency analysis.

It seems hardly necessary to add that the procedure used in this experiment does not reproduce the rich complexity of judgements about people in ‘real life’. Nor is it intended to do this. The study is an attempt to analyse some of the simpler aspects of the ‘context’ within which we come to attribute one or another characteristic to an individual. It is also an expression of the psychologist’s conviction that human behaviour is a natural phenomenon – and there are uniformities in natural phenomena. Therefore, he cannot rest contented with the assumption that because things are complex they must also be ineffable and unanalysable, or that terms such as ‘intuition’ mark the end of enquiry. Much of human behaviour is the result of swift and highly efficient computations of probabilities – of the likelihood that certain events tend or do not tend to happen together. The study by Bruner and his colleagues shows that this is so in the ‘perception’ of people; but much experimental ingenuity and imagination will be needed in order to gain an even partial understanding of this activity though normally it appears as one only of our simple daily routines.

Conclusion

As most such surveys, the present one cannot claim to be complete or unbiased. It is hoped, however, that it will have succeeded in convincing the reader (i) that the author was justified in not attempting at the outset a 'definition' of social perception; (ii) that though parts of what is called 'social perception' are not quite 'social', and other parts certainly not 'perceptual', this is an important area of problems in which only a beginning has been made; and (iii) that the promised land lies in the direction of a close integration of most problems in social perception with concepts and findings in general cognitive theory.

There are two justifications for the use of the wider term 'cognitive' instead of 'perceptual'. One has already been stated: even a wide and tolerant definition of perception would be strained by the inclusion of some of the investigations referred to above. The second is related to the first: inferring intelligence from height of forehead is inferring, not 'perceiving'. Deciding that a man is probably industrious 'because' he is a German, or that he is likely to have a sense of humour 'because' he is intelligent and has a twinkle in his eye, is certainly inferring, and not perceiving. Most phenomena in social perception present problems in inference, categorization and judgement. Though there can be no sharp distinction between these and more conventionally perceptual problems – since categorization, inference and judgement have come to play a prominent role in our understanding of perception – the fact remains that we must turn to the largely unexplored and promising relationships between the field of social perception and those of thinking, learning, and judgement.

Acknowledgements. This chapter was written during a year's visit to the United States. I am deeply grateful to a number of friends, too many to enumerate here, for the opportunities they gave me to discuss their work with them; also to Mrs Eleanor Horan from the Laboratory of Social Relations at Harvard University for her patient typing of the first draft.

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., and SANFORD, R. N. (1950). *The Authoritarian Personality*. New York: Harper.
- ALLPORT, G. W., and PETTIGREW, T. F. (1957). 'Cultural influence on the perception of movement; the trapezoidal illusion among Zulus'. *J. abnorm. soc. Psychol.*, **55**, 104-13.
- ASCH, S. E. (1946). 'Forming impressions of personality'. *J. abnorm. soc. Psychol.*, **41**, 258-90.
- ASCH, S. E. (1952). *Social Psychology*. New York: Prentice Hall.
- BROWN, R. W., and LENNEBERG, E. H. (1954). 'A study in language and cognition'. *J. abnorm. soc. Psychol.*, **49**, 454-62.
- BRUNER, J. S. (1957). 'Going beyond the information given'. In Bruner, J. S., *et al.*: *Contemporary Approaches to Cognition*. Cambridge: Harvard Univ. Press, 41-70.
- BRUNER, J. S., SHAPIRO, D., and TAGIURI, R. (1958). 'The meaning of traits in isolation and in combination'. In Tagiuri, R., and Petrullo, L. (eds.): *Person Perception and Interpersonal Behaviour*. Stanford: Stanford Univ. Press, 277-88.
- BRUNSWIK, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley: Univ. of California Press.
- COHEN, J. (1959). 'Review of "Person perception and interpersonal behaviour"'. *Brit. J. Psychol.*, **50**, 82-3.
- DIXON, N. F. (1958). 'Apparent changes in the visual threshold as a function of subliminal stimulation'. *Quart. J. exp. Psychol.*, **10**, 211-19.
- GILBERT, G. M. (1951). 'Stereotype persistence and change among college students'. *J. abnorm. soc. Psychol.*, **46**, 245-54.
- HASTORF, A. H., RICHARDSON, S. A., and DORNBUSCH, S. M. (1958). 'The problem of relevance in the study of person perception'. In Tagiuri, R., and Petrullo, L. (eds.): *Person Perception and Interpersonal Behaviour*. Stanford: Stanford Univ. Press, 54-62.
- HOCHBERG, J. (1957). 'Psychophysics and stereotypy in social perception'. In Sherif, M., and Wilson, M. O. (eds.): *Emerging*

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

Problems in Social Psychology. Norman: Univ. of Oklahoma Press, 117-41.

- KATZ, D., and BRALY, K. W. (1933). 'Racial stereotypes of 100 college students'. *J. abnorm. soc. Psychol.*, **28**, 280-90.
- LAMBERT, W. E., HODGSON, R. C., GARDNER, R. C., and FILLENBAUM, S. (1960). 'Evaluational reactions to spoken languages'. *J. abnorm. soc. Psychol.*, **60**, 44-52.
- LENNEBERG, E. H. (1957). 'A probabilistic approach to language learning'. *Beh. Sci.*, **2**, 1-12.
- LENNEBERG, E. H., and ROBERTS, J. M. (1956). 'The language of experience: a study in methodology'. *Suppl. Internat. J. Amer. Linguist.*, **22**, No. 2.
- MACLEOD, R. B. (1951). 'The role of phenomenological analysis in social psychological theory'. In Rohrer, J. H., and Sherif, M. (eds.): *Social Psychology at the Crossroads*. New York: Harper, 215-241.
- MCCURDY, H. G. (1956). 'Coin perception studies and the concept of schemata'. *Psychol. Rev.*, **63**, 160-8.
- MILLER, A. (1945). *Focus*. New York: Harcourt, Brace.
- RAZRAN, G. (1950). 'Ethnic dislikes and stereotypes: a laboratory study'. *J. abnorm. soc. Psychol.*, **45**, 7-27.
- RIVERS, W. H. R. (1904). 'Observations on the senses of the Todas'. *Brit. J. Psychol.*, **1**, 321-96.
- SECORD, P. F., BEVAN, W., and KATZ, B. (1956). 'The Negro, stereotype and perceptual accentuation'. *J. abnorm. soc. Psychol.*, **53**, 309-14.
- SINGER, B. R. (1956). 'An experimental inquiry into the concept of perceptual defence'. *Brit. J. Psychol.*, **47**, 298-311.
- TAJFEL, H. (1957). 'Value and the perceptual judgement of magnitude'. *Psychol. Rev.*, **64**, 192-204.
- TAJFEL, H. (1959a). 'Quantitative judgement in social perception'. *Brit. J. Psychol.*, **50**, 16-29.
- TAJFEL, H. (1959b). 'A note on Lambert's "Evaluational reactions to spoken languages"'. *Canad. J. Psychol.*, **13**, 86-92.
- WISHNER, J. (1960). 'Reanalysis of "Impressions of personality"'. *Psychol. Rev.*, **67**, 96-112.

Chapter 2

MOTIVATION AND CONFLICT

PETER ROBINSON AND MICHAEL ARGYLE

Oxford University

Human needs and drives have been found difficult to study both because there is no agreed way of classifying them and because they are hard to measure. Animal needs and drives are easier: animals can be deprived or injected and their subsequent efficiency of learning and performance compared. The indices commonly used do not necessarily show a high degree of agreement, but at least the units of measurement can be clearly defined.

To a very limited extent it is possible to proceed in the same way with human beings – several experiments have been done with food deprivation. In the case of drives whose physiological basis is more obscure, such as the needs for achievement (nAch) and affiliation (nAff), investigators have used analogous techniques. It was assumed that nAch can be aroused by artificial failure at a task, but success had similar effects on the measurements made! Such drives can also be aroused by talking to subjects about the area of behaviour in question.

Various ways of measuring the strength of particular human drives have been suggested. These can be validated either in terms of their sensitivity to arousal, as described above, or by their correlation with indices of overt behaviour of different people. Three types of measure are used in the experiments to be described here – projection tests, experimental situations, and questionnaires.

Projection tests such as the Rorschach and TAT have not fared too well in validation studies, but parts of the TAT have had a new lease of life in the assessment of nAch and other drives. McClelland (1951,

1953), Atkinson (1958) and their associates have carried out a very important series of investigations in which drives are measured from stories written to selected TAT cards. The basic finding is that stories written by subjects whose drive-state has been aroused contain more imagery related to the drive in question. An empirical demonstration of this had been made with the hunger drive a decade before: Sanford (1936) compared the stories which two groups of young children told about some pictures, one group telling the stories when hungry, the other group telling them just after a meal. He found that the former group gave more 'food responses'. Atkinson and McClelland (1948) followed this lead and deprived three groups of subjects of food for 1, 4, and 16 hours and then had them write stories about some pictures. They found that certain categories of thought increased with increasing deprivation: themes concerned with food deprivation, characters expressing a need for food and activity successful in overcoming deprivation. There was a corresponding decrease in the amount of goal activity itself, i.e. eating. A further experiment by McClelland, Clark, Roby, and Atkinson (1949) showed that a comparable method could be applied to the concept of nAchievement. In this case instead of using a deprivation method, tests given before the storywriting were introduced as being capable of selecting those persons with organizational and administrative ability. The instructions in the arousal groups were assumed to give rise to a temporary state of motivational arousal. The further inference was made that the categories of thought differentiating the aroused from the non-aroused group could provide a possible index of a more stable motive – that differential use of the categories by persons reflected individual differences in motivation. Some evidence has been obtained that, for example, over-achievers at college (in relation to their ability) score high on nAch measured in this way (McClelland *et al.*, 1953). These scores also have interesting relations with socialization experiences and sociological variables.

Experimental situations have been used to measure motivation, particularly what is known as the *level of aspiration*. A subject performs a task where the performance can be expressed as a number. After each trial he is told his score and asked what he aims to get next

time. The difference between the score aimed for and the last score obtained is found to be relatively constant for each person in a given task, and this difference is known as the discrepancy-score (D-score). This variable was intensively studied some time ago; the earlier literature is summarized by Lewin and others (1944).

Questionnaire methods have not been widely used in this field, partly because they tend to miss the unconscious elements of motivation. However, indirect verbal methods have been used in studies of conflicts within the personality, such as between the self, the ego-ideal and the super-ego. Cronbach (1953) reviews work using the Q-sort method in which subjects are asked to rank adjectives in the order in which these apply to themselves, their ideal self and so forth. A more recent development in psychological measurement is the Semantic Differential of Osgood and his co-workers (1957), which is described below and used in the third experiment. There is still bound to be some doubt over how far unconscious drives are being tapped, and this point is discussed in connection with this experiment.

The three experiments presented in this chapter are concerned with basic theoretical problems of human motivation, which can only be regarded as partly solved at the present time. *The first experiment* starts by repeating an established result – the influence of achievement arousal on TAT stories. It goes on to study the relation between nAch and level of aspiration: this is a complex relation, and various theoretical problems are raised by it, such as the possible separation of nAch into the hope of success and the fear of failure.

The second experiment is concerned with the arousal and measurement of sexual needs in an analogous way to nAch. It is found that arousal produces disguised imagery of a Freudian type, but no overt sexual imagery. When inhibiting forces are reduced by means of alcohol, overt imagery appears. This experiment makes an interesting link with psychoanalytic theory, and shows the importance of internal restraints in the expression of motivation.

The third experiment follows up this idea of conflict between drives and restraining forces. Self-ratings are obtained on guilt-feelings and the total guilt score is related to the discrepancy between self and super-ego. The origins of the super-ego are studied by comparing the

semantic descriptions of the super-ego with 'father's demands', 'father's behaviour', and so on.

ACHIEVEMENT AND LEVEL OF ASPIRATION

The empirical problem. It was stated in the introduction that instructions could be used to change the content of imaginative stories written in response to pictures in such a way that certain achievement-related categories of thought increased in frequency. These increases were thought to be the result of an increase in motivational arousal and it was further suggested that the incidence of these categories could be used to show up individual differences in achievement.

Our first aim will be to show that the content of imaginative stories can be changed significantly (Prediction 1) with respect to the quantity of achievement imagery present. This demonstration is basic to all the subsequent work done in this field. Psychologists with clinical experience are aware of the large variety of sources for imaginative stories in the Thematic Apperception Test (Murray, 1938), and may doubt that motivational factors can be separated out without corroborative evidence. Persons not conversant with the technique are sceptical about obtaining measures of motivation from the sort of stories people write.

The second aim is to relate achievement scores to level of aspiration in a simple motor task which has few similarities to story-writing. Atkinson (1957, 1958) has constructed a theoretical model relating these variables. In any task the level of difficulty selected as a goal is the product of motive intensity, incentive value and the expectancy of its attainment. These three variables apply to two aspects of motivation - 'motivation to achieve' and 'motivation to avoid failure'. The goal selected will be the one for which the summated motivation is most positive. Measures of incentive value and expectancy can be derived from subjective probabilities of success and failure. If 'motivation to achieve' is greater than 'motivation to avoid failure' the resultant motivation will be most positive for tasks of intermediate difficulty (Atkinson, 1957, see Table 1). If 'motivation to avoid

failure' is greater than 'motivation to achieve', the person should leave the situation, but if forced by various constraints to state a level of aspiration he should select tasks which minimize anxiety to avoid failure. He should select either a very hard or a very easy task. Which course is in fact adopted will be a function of extraneous factors and we assume that these factors are randomly distributed in any given sample of subjects.

Atkinson further adduces evidence to show that low nAch scores go with relatively high 'motive to avoid failure' while high nAch scores indicate relatively high 'motive to achieve'. We can therefore predict that the variation in D-scores of low nAch subjects will be greater than that of high nAch subjects (Prediction 2).

Method

(1) *Subjects*. A group of twenty-four male students should be sufficient to give significant differences on the hypotheses we are to test. As is usual in psychological experiments, it is best if the sample is homogeneous on relevant factors, such as religious denomination, social class, and sex in this case (Atkinson, 1958).

(2) *Measurement of dependent variables*. For the assessment of nAchievement, two sets of four pictures are recommended. Sources are given in the general list of references at the end.

Set A. (1) Man seated at drawing board (ADI, 28), (2) The inventors (McClelland *et al.*, 1953), (3) Young boy watching aircraft taking off (ADI, 25), (4) 'Homeland' (ADI, 24).

Set B. (1) Boy in checked shirt at a desk, an open book in front of him (McClelland *et al.*, 1953), (2) Elderly man reading what he has typed (ADI, 48), (3) Boy with operation scene in background (Murray, 1943, T.A.T. 8BM), (4) Man working on papers at office desk (ADI, 9).

These two sets have been found to be satisfactory with English university students and student teachers. The two sets A and B have been found to elicit about the same amount of achievement imagery, so that they can be used safely as before and after measures.

For the level of aspiration task each subject requires sheets of

paper containing five rows of ten circles each about 1 cm. in diameter. Below this matrix should be boxes opposite the instructions 'No. of faces completed', and 'Goal for next set'. Ten such sets are needed for each subject. Stopwatches are required for the experimenters.

(3) *Procedure.* The subjects should be assigned at random to two groups of equal size. These two groups undergo exactly the same treatment except that the sets of pictures are given in reverse order.

The following questions should be written on the blackboard:

- (1) What is happening? Who are the persons?
- (2) What has led up to the situation?
- (3) What is being thought or said? What is wanted?
- (4) What will happen?

The first set of stories are supposed to be written under relaxed conditions and if the group of subjects seems to be apprehensive it is quite in order to preface the instructions with some reassuring remarks.

The instructions for the relaxed conditions are as follows:

The main part of what we are going to do this afternoon requires you to answer these questions on the blackboard about some pictures that I have here. It is reasonable that you should have some practice first. The sort of story people write about a picture depends on the sort of picture shown. A fellow worker of mine at X . . . university is trying to find out what sort of stories are written to these four pictures. He is only trying to make a large collection and is not interested in the people who write them. Your stories will add to his collection and at the same time give you practice. There is no need for you to write your names on the papers.

This is what will happen. I will show you the pictures for about fifteen seconds, and you will then have four and a half minutes in which to write a story about the picture. The questions on the board are a guide. If you answer each one of them you should produce a short story with a beginning, a middle, and an end. There are no right or wrong answers, and please just don't describe the picture. Write a story. This practice will give you an opportunity to judge how long you have to complete each story and get you

into the swing of the procedure. You can just write down the first story that comes into your mind when you see the picture. Here is the first one.

The picture is then presented without any verbal title and after four and a half minutes the subjects are asked to stop writing and the next picture is shown. If questions are asked concerning the aims of the investigation, replies should be as non-committal as possible, except in so far as they are related to experimental procedure. The subjects must be given no clue as to how the stories will be evaluated and it should be emphasized that the situation is a free one and subjects can write about anything.

For the level of aspiration task, it is preferable to break up the subjects into small groups of four or five persons, the fewer the better. Each matrix set should be numbered and names should be added to the first sheet.

The instructions for the motor task are:

What you are now going to do is a simple task. You have sets of fifty circles in front of you. The object is to draw four lines in these circles producing a face. You must draw the lines in this sequence (Experimenter demonstrates), working from left to right across the rows and then on to the next line. You have thirty seconds for each set. At the end of thirty seconds I will tell you to stop. You count the number of completed faces and enter this number in the appropriate box. In the box below you enter the number of faces you aim to complete in the next thirty-second period. This number should represent your immediate goal for the next run, it should express an intention, not a vague hope. It is essential that no one gains any information about the performances of anyone else, so please don't mention anything to do with this task until you have finished *everything* here this afternoon.

Nine trials should be run with only a slight pause between runs. The experimenter should not say anything to the subjects except 'Are you ready? Go' and 'Stop'. He should not make any reaction to the subjects' performance. If a subject asks whether he is meant to fill in as many circles as he can, the experimenter should reply that he

can do what he likes. The faces should be constructed in the order right eye, left eye, nose, mouth, with just a dash for each feature. Ten matrices are necessary because the first trial is ignored in the scoring and the extra sheet at the end ensures that the subject will state a goal on the ninth sheet.

On completion of this task the subjects are brought back into their groups again. The following instructions are then read by a different experimenter.

The task you have just completed was taken from a group of tests which is designed to give an assessment of a person's organizational abilities. This particular task is related to perseverance, but unfortunately I have not time now to explain the scoring details. Now there are some more pictures. Unlike the others, these have been used extensively in the United States and in this country. This is a test of what psychologists call 'creative imagination'. Here again there are no right or wrong answers and so you are free to make up any kind of story that you like. Don't just describe the pictures. Tell an interesting story using these questions as a guide. Work sufficiently fast in order to finish in time. As before you will have four and a half minutes for each story. This time will you write your names on the papers, please? Here is the first picture. The procedure is as before.

Statistical treatment and results

Only some of the criteria for scoring nAchievement will be given here; if the information is insufficient, reference should be made to one of the works already cited.* The most vital decision to be taken is whether a story contains any achievement imagery (AI). Two conditions must be fulfilled before AI can be scored: (1) Some activity must be evaluated. Adjectives such as 'good' and 'bad' or adverbs such as 'well' often provide the necessary justification for moving on to the second condition. (2) A character in the story must be emotionally or affectively concerned about his performance in relation to this standard. 'Wanting' and 'trying' reveal such personal involvement.

* McClelland *et al.*, 1953, or Atkinson, 1958, chap. 12.

If these two criteria are met, e.g. He wants to be a good doctor, the story is scored +1. If there is any doubt the story should be scored 0 for doubtful imagery (TI) and if it is clear that the story has no achievement imagery it is scored -1. In both these cases the story is not scored further.

These strict conditions have been relaxed in American work (McClelland *et al.*, 1953) and references to long term career goals and unique accomplishments such as inventions qualify as achievement imagery. An additional requirement is that no ulterior reason must be appended to a statement containing achievement imagery, e.g. 'He wants to be a good doctor in order to earn a lot of money and live in idle luxury'. A useful rule is not to score if in doubt.

If achievement imagery is present the story is now analysed for the following categories:

(a) Need (N): An explicit statement of a desire to reach an achievement-related goal.

(b) Instrumental activity (I): An explicit indication of the means taken to reach such a goal.

(c) Anticipatory goal-states (G_+ and G_-): G_+ is scored when a character in the story looks forward to the successful completion of an achievement-related activity and anticipates the pleasure to be derived from his success. Anticipation of failure or worrying about the outcome scores G_- .

(d) Affective states (G_+ and G_-): If the character *actually* derives pleasure from success G_+ is scored and if he is sad at failing G_- .

(e) Blocks (Bp and Bw): These are obstacles preventing someone from reaching an achievement-related goal. Bp refers to a personal handicap or misfortune, Bw is a block produced by another person or the environment.

(f) Nurturant Press (Nup): When someone is helped by another person towards an achievement-related goal, Nup is scored.

(g) Achievement thema (Ach th): If the achievement-related activities constitute the whole plot of the story then Ach th is scored.

Each category mentioned scores +1, giving a possible score of +11 (AI, N, I, G_+ , G_- , G_+ , G_- , Bp, Bw, Nup, and Ach th),

and no category can be scored more than once. To obtain a person's total score the four picture totals are summated.

Motor task: The first trial is omitted from the analysis, leaving eight. Two indices are required from these trials. The first is the discrepancy score (D-score) which is here defined as the number of faces completed on any given trial subtracted from the number of faces aimed at for the *next* set. Negative signs are given to D-scores where the goal is less than the previous run. Mean D-scores are required for the first four trials, the second four, and for all eight trials. Secondly, the number of times an undertaken goal is reached or exceeded is assessed. This will be called NTR.

Reliability: Before testing the predictions made, we should assess the reliability of our measures. The scoring of nAchievement may be checked by having two persons score each subject's stories and measuring the extent of agreement with a rank order correlation. Within the present framework it is not possible to test the reliability of the nAchievement scores themselves. The split-half method can be used to measure the reliability of the D-scores, correlating the first four with the second four. The question of interjudge reliability does not arise.

Testing the predictions

Prediction 1: The interpolation of the motor task and the intervening instructions should produce a significant increase in nAchievement scores from the relaxed to the achievement-oriented conditions. The significance of the difference between the two conditions can be most adequately assessed by means of the non-parametric Wilcoxon Matched-Pairs Signed-Ranks Test for two related samples (Siegel, 1956). The appropriate statistic should be calculated for each group.

Prediction 2: The variance of the D-scores of the low nAchievement subjects will be greater than the variance of the high nAchievement subjects.* This may be tested in three ways.

- (a) If each group of subjects is divided at the median with respect

* This is because low nAch Scores should have a high fear of failure, and are very high or very low to minimize the product of incentive value and failure expectancy.

to its achievement-oriented (aroused) nAchievement scores, an F test can be used to compare the variance of the low and high groups.

(b) Alternatively the subjects can be divided on the basis of their D-scores into three groups, a high, middle, and low group, such that at least half of the subjects are in the middle group. It has been found that +1.00 and 2.25 have been useful breaking points. We can now predict that the middle group should have higher nAchievement scores than either the high or the low groups and the two combined. The significance of the differences of the aroused nAchievement scores of the three groups can be calculated using the Mann-Whitney U test (Siegel, 1956).

(c) The subjective probability of success is measured indirectly from the D-score, but can be measured directly by the NTR index. This provides an index of the subject's actual probability of reaching his stated goal. The maximum score is 6, the minimum 0, and we would expect subjects with relatively high 'motivation to avoid failure' to approach these two limits. An important feature of the high D-score strategy is that the goal should never be reached. Secondly, many subjects will have built up a tendency to take higher risks than Atkinson's simplest version of his theory mentions. He elaborates that a history of success will allow a person to tolerate higher odds against success. Accordingly we can classify subjects with NTR 0 together, subjects with NTR 4, 5, and 6 together, and subjects with NTR 1, 2, and 3. We can then use the Mann-Whitney U test in the way described under 2.

Discussion. The fact that more achievement imagery appears after subjects have been aroused has been the starting-point for all the research into achievement motivation. It will be seen in the experiment to be described next that sexual arousal does *not* increase the overtly sexual content of the stories written. Freudian theorizing about this drive also suggests another theoretical model to account for 'motivated imagery': it may be a substitute form of satisfaction for the drive. Atkinson and McClelland assume a rather different model: when a person is activated by a certain drive this manifests itself in all areas of his behaviour, including imagery. If the first model is correct, then people with much achievement imagery should be

less successful in real life: if the second model is correct, those with achievement imagery should strive harder in real life. On the whole the second is better supported by the research findings – for example the higher nAch scores of over-achievers in school. In the case of aggression, it has been found that for boys whose mothers encouraged aggression, overt and fantasy aggression correlated $\cdot 43$, but for boys whose mothers discouraged it, they correlated $-.41$. Since achievement is encouraged, sex often discouraged, by parents, it is reasonable that different relationships should hold for the two drives, and that different models concerning the production of motivated imagery should be relevant.

Summary. Subjects write stories about certain TAT cards under relaxed conditions, carry out a level of aspiration task and then write more TAT stories while aroused for achievement. It is found that the second set of stories contains more achievement imagery than the first, and that subjects with low achievement need, as estimated from the imagery, have either very high or very low levels of aspiration.

SEXUAL MOTIVATION AND CONFLICT

Introduction. Preliminary research by members of the McClelland-Atkinson group showed that attempts to arouse sexual motivation in subjects did not produce TAT stories with an overt sexual content. Meanwhile intensive case studies showed that some subjects highly motivated to achieve did not express this in the TAT, as a result of anxiety over achievement. It seemed possible that inhibitory forces in the personality were preventing the expression of sexual motivation in the TAT. If this is so, overt sexual imagery should be produced following sexual arousal if anxiety can be reduced in some way. There is considerable experimental and other evidence that alcohol reduces anxiety, and this is the method used in the experiment to be described. Clark (1952) used various experimental groups which were subjected to various forms of sexual arousal, some in an alcoholic state, some not. The expectation is that the sexually-aroused, alcoholic group will show more overt sexual fantasy in the TAT than groups which were not aroused, or not alcoholic. Clark used 35–40 subjects in each con-

dition and found large and significant differences in the expected direction.

Freud distinguishes the manifest from the latent content of dreams and other imaginative productions. He states that the sexual instinct can be given disguised expression through the symbolism of the latent content. It might be expected in the present experiment that such disguised expression would occur in the sexually-aroused, non-alcoholic group, since anxiety is still operating to prevent direct expression via the manifest content. Clark and Sensibar (1955) tested this hypothesis by re-analysing the data of the earlier experiment described, and found more sexual symbolism for the aroused non-alcoholic group than for either of the others.

The aim of the present experiment is to confirm two of Clark and Sensibar's findings:

(a) The sexually-aroused, alcoholic group will show more overt sexual imagery in the TAT than the sexually-aroused non-alcoholic group or the control group.

(b) The sexually-aroused, non-alcoholic group will show more symbolic sexual imagery in the TAT than the sexually-aroused, alcoholic group, or the control group.

The version to be described is rather better controlled than the original in that Clark used a fraternity beer-party for his alcoholic condition, whereas in our version the alcohol is administered under laboratory conditions. Clearly a beer-party setting introduces variables other than alcohol, so that it is not certain that this is the relevant variable.

Method

(1) *Manipulation of experimental variables.* Sexual arousal can be brought about in a variety of ways, though the requirements of decorum and public relations place some limitation on these. The following has been found to work with male students. The aroused group are placed in a room with an attractive female experimenter, who administers the TAT. If possible the timing should be arranged so that both aroused groups have the same female experimenter, in

order that differences in arousal power are eliminated; the non-aroused group have a male experimenter. Prior to this, the aroused group are shown a number of magazine photographs of attractive girls and asked to make paired comparisons of their physical attractiveness. This can be presented either as an exercise in paired comparison, or as a study of the degree of consensus on the relative attractiveness of different types of somatotype.

The alcoholic condition was produced in Clark's experiment by testing the subjects while they were at a beer-party. As argued above, this entails other uncontrolled changes. In the laboratory the subjects can be given three glasses of sherry or two pints of beer; this produces the desired results while not rendering the subjects uncontrollable.

(2) *Measurement of dependent variables.* Five cards from the TAT are used, selected for their likelihood of producing sexual imagery. Cards 4, 6 GF 10, 12 BG, and 13 MF are suggested, though there are other possibilities. The following instructions are given to the subjects:

We are now going to write some short stories, the themes of which are to be inspired by some pictures I have here. You will have a minute to look at each picture, and five minutes to write a story suggested by it. This is a test of creative imagination: don't just describe the picture, tell a story with a beginning, middle, and an end, and say what the characters are feeling and thinking. Are you ready?

Overt sexual fantasy can be assessed in three categories:

- (a) reference is made to the characters having sexual intercourse +3
- (b) evidence of kissing, dancing or other physical contact +2
- (c) characters seen as in love, married etc., but without physical contact +1.

Each of the five stories is scored from 3 to 0 and a total obtained for each subject. If time permits, more than one person should score each set of stories, to obtain a more reliable average, and to give evidence about the reliability of scoring.

Symbolic sexual fantasy can be assessed in two categories:

- (a) the central theme of the story involves some activity symbolic

of sexual activity, such as climbing through a window, flying, etc., +2

- (b) mention of classical Freudian symbols, for the genitals, breasts etc., provided the objects do not actually appear in the picture +1.

This is harder to score than the manifest fantasy, and some general discussion in the class may help to decide on the scoring of particular stories. Again a total score is worked out for each subject. These scores can be kept quite anonymous, since it is only the scores for members of each experimental group that are required.

(3) *Procedure.* In order to run this smoothly as a class experiment, careful planning and staffwork are necessary. Several assistant experimenters, drawn if possible from the class, must be briefed beforehand on their duties.

When the group is first assembled members are allocated at random to each of the three conditions. Since the experiment is designed for male subjects, any women in the class must either be experimenters, or possibly be included in the control group.* A separate room is used for each condition, and an experimenter is in charge of each group. The person in charge can move round making sure that all goes according to plan. If only one set of the TAT is available, things must be staggered sufficiently for the cards to be passed on from group to group.

When each group has finished writing the TAT stories, it should assemble in a central room. The point and design of the experiment are then explained, and instruction given in scoring the stories.

Statistical treatment and results

	CONTROL	AROUSED, NON-ALC.	AROUSED ALC.
OVERT	A	B	C
SYMBOLIC	D	E	F

* If there are more women than men in the class, the women would have to be the subjects.

In each box are listed the scores of the various subjects in that condition. The data can be initially examined by simply working out the average scores in each cell, to see if the hypotheses are going to be confirmed. It will be remembered that C should be greater than A or B, and that E should be greater than D or F. To test the significance of these differences, the Mann-Whitney test is the most convenient. Greater sensitivity with a small number of subjects is obtained if C is compared with A and B together for the first hypothesis, E with D and F together for the second.

Discussion. This experiment is of interest in a number of ways. From the point of view of measuring motivation, it supports the idea that TAT stories reflect motivational states, since both of the aroused conditions produced more sexual imagery—though of different kinds. In this experiment sexual motivation is aroused, and the only outlet provided is in fantasy. It may not follow that individuals with high scores have high sex drives, or that they engage in much sexual activity, but simply that they cannot find an outlet for their sexual needs, as was pointed out in connection with the first experiment.

The difference between the alcoholic and non-alcoholic condition supports the framework of ideas presented at the beginning. It seems reasonable to conclude that overt sexual fantasy is inhibited by anxiety, and that alcohol reduces this anxiety. Very possibly other drives as measured by the TAT method are not being fully expressed for the same reason. In this case a truer measure of, say, need for achievement would be obtained by adding to the overt achievement imagery the amount of symbolic achievement imagery. However, we do not yet know what the symbols for achievement *are*. This suggests another line of research—doing for other drives what Freud did for sex, and finding out what their disguised symbolic expressions are. This could be done by comparing the stories of aroused groups under alcoholic and non-alcoholic conditions.

Summary. Subjects are divided into three groups. The first group is simply given 5 TAT cards, and their stories are scored for overt and symbolic sexual imagery. The second group have the level of their sex-drive temporarily increased before being given the cards.

The third group are in addition given a certain amount of alcohol before being shown the cards.

It is found that the third group scores highest on overt sexual imagery, while the second group scores highest on symbolic sexual imagery.

SUPER-EGO CONFLICT AND GUILT

Introduction. It is often supposed that guilt feelings arise from a failure to live up to the demands of conscience. This hypothesis was tested by Brahmachari (1954) by a questionnaire method. He drew up a list of forty-two items of the type 'Is it desirable to abstain from alcoholic drink?' and presented it to subjects with three sets of instructions. Firstly they were asked how far they agreed or disagreed with the item; secondly they were asked how far they realized this ideal in their own actual conduct; thirdly they were asked how far any discrepancy between the first two answers gave rise to worry or conflict. The first two answers were recorded on scales from +5 to -5, the third from 0 to 5.

For the 120 subjects used by Brahmachari, the correlation between discrepancy and conflict experienced was only .16 (not quite significant at the 5 per cent. level). One possible explanation of this very low correlation is that moral attitudes and conduct are being enquired about in too direct a way - a person might misreport his actual conduct or be unconscious of the real demands of his conscience on such issues. The present experiment introduces a less direct measure of self/super-ego conflict, based on Osgood's semantic differential (1957).

In this test, it is possible to assess the meaning of a concept to a subject by asking him to rate it on a series of 7-point scales, such as:

strong — : — : — : — : — : — : weak

Extensive statistical analyses have been made of the relation between different scales. For personality research four main factors of meaning have been found - evaluation, activity, potency and uniqueness. A number of scales are known to be the best measures of each factor.

Appendix I shows the set of scales we have been using: it consists of several for each dimension in random order. To administer the test, a concept is given at the head of each sheet of scales and the subject rates each concept against all the scales. Two particular concepts are of interest here, 'THE KIND OF PERSON I ACTUALLY AM' and 'THE KIND OF PERSON I OUGHT TO BE'. These are used to measure the meaning of self and super-ego. Other concepts which can be used are 'THE KIND OF PERSON MY FRIENDS WOULD LIKE ME TO BE' and so forth. The present concern is to measure the 'distance' between self and super-ego. This is done by scoring each scale marking from 1 to 7, subtracting the ratings for self and super-ego for corresponding scales and finding total differences. (A root-mean square has been suggested, but is more laborious, while the present method gives closely correlated results, see Osgood *et al.*, 1957, p. 90 f.)

In order to measure guilt feelings, a development of Brahmachari's scale is used. In the original experiment an item analysis of the moral attitudes scale was carried out – showing which items constitute a unitary dimension of acts thought morally desirable. We have used simplified versions of these remaining items as a basis for a guilt scale. This has been purified by factor analysis, the first factor accounting for 38 per cent. of the variance. Average scores are worked out for each subject.

We can also enquire about the origins of super-ego demands, and this can also be studied by means of the semantic differential. If it is found that 'THE KIND OF PERSON I OUGHT TO BE' is scored very much the same as 'THE KIND OF PERSON MY FATHER THINKS I OUGHT TO BE', this suggests that father's demands may be part of the origin of the super-ego. Our investigations have shown that the super-ego is described as being very similar to the demands of various other people, not to their behaviour; for school-children the super-ego is nearest to the demands of the same-sexed parent, for students it is nearest to the demands of their friends, with large individual differences. Blum (1949) found with the Blacky test that the super-ego of American students was based on the same-sexed parent, but he did not study the influence of friends.

Aim of the present experiment. The present experiment combines an extension of the Brahmachari experiment (using an indirect measure of super-ego conflict) with a study of the origins of the super-ego.

Two hypotheses are being tested:

- (i) There will be a positive correlation between the extent of self/super-ego conflict and the guilt feelings of individuals.
- (ii) The super-ego will be described as nearer to friends' demands than to parental demands, or to the reported behaviour of either.

Method

(1) *Measuring the personality dimensions to be correlated.* Self/super-ego conflict is measured by the indirect method described above. The set of scales used is given in Appendix I. These are based on research by Osgood *et al.* (1957) and on unpublished work by the same group. The only measure used here is the 'distance' between the self and super-ego obtained by the average of the differences on the individual scales. It is also possible to obtain a measure of super-ego strength from this test. On the 'evaluation' dimension, the mean score for *happy, honest, kind, clean and fair* can be computed. For the 'activity' dimension, average *excitable, active, impulsive and fast*; for the 'potency' dimension, average *strong, large, hard, rugged and rigid*. Guilt can be measured by a scale given in Appendix II, whose construction has been described earlier.

(2) *Measuring correlates of the super-ego.* A series of semantic scales is prepared, covering the demands (and if time, the behaviour) of each parent, and 'MY FRIENDS'. The semantic distances are calculated for each subject between his super-ego and the possible origins, by the method used before.

(3) *Procedure.* When the group has assembled, the two personality tests are given out together with the necessary instructions for the semantic differential. No explanation is given at this stage about what the experiment is about, or what the tests are supposed to measure.

When all have completed these tests, the scoring system is explained, and subjects mark their own tests. It has been found that considerable interest is aroused by this experiment if subjects progressively discover their own scores on these dimensions, the relation of these to the scores of others, and the way their own scores contribute to the verification of hypotheses.

Statistical treatment and results. The first hypothesis can be examined graphically to begin with. Super-ego conflict and guilt can be plotted on the blackboard, and individual scores plotted.

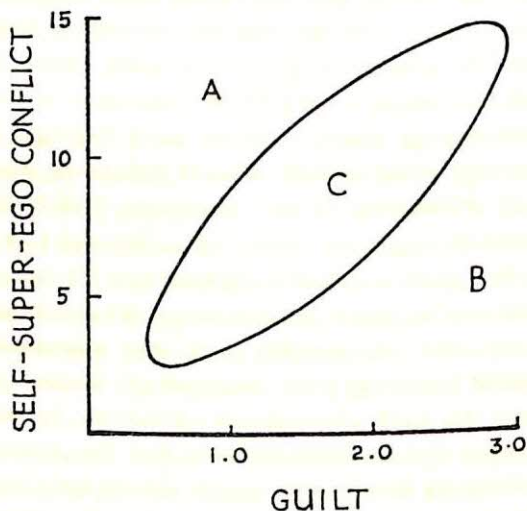


FIG. 1.

If there is a very close correlation between guilt and super-ego discrepancy, the scores would be expected to lie in an area such as C in Fig. 1. It will be found, however, that the correlation is not so close as this, and that some subjects have too little guilt for a given discrepancy (area A), others have too much guilt (area B). The extent of the correlation between the two variables can be worked out by means of a rank-order correlation, such as Spearman's rho, and this will show the extent to which the hypothesis is confirmed.

We have found that there is a rather higher correlation between guilt and self/super-ego discrepancy on the evaluative factor alone.

This is the average discrepancy score for the items measuring the 'evaluative' dimension, which were listed above.

The second hypothesis is tested by finding the distances between super-ego and demands of each parent and of friends. A simple arithmetical average shows which of these the super-ego is closest to. The significance of the difference between the distances between SE/father's demands and SE/mother's demands, for instance, can be found by the Wilcoxon test.

Discussion. The finding that there is a positive correlation between super-ego conflict and guilt gives some support to the usual ideas on this subject. The fact that the correlation is higher for discrepancy on the evaluative dimension alone shows that although people think they ought to do a lot of things they don't do, it is only certain shortcomings which produce guilt feelings. This may be because it is only some of these areas of behaviour which have been socialized by withdrawal of love discipline, widely found to be a source of guilt feelings (e.g. Sears, Maccoby, and Levin, 1957).

Brahmachari gives a different explanation. He examined the discrepancy and conflict scores for different items and found for example that men reported considerable guilt over unconventional sexual activities while reporting little discrepancy; women had more discrepancy and less guilt. He suggests – following Freud (1923) – that part of the super-ego is unconscious and that this creates guilt feelings on certain items on which there is little discrepancy between the self and the conscious part of the super-ego. The fact that the *inter-item* correlation for Brahmachari's subjects between discrepancy and guilt was .56 and the *inter-subject* correlation only .16, suggests that individuals rather than items vary in the unconsciousness of the super-ego. The next step might be to use projection test measures both of discrepancy and of guilt to see if higher correlations can be obtained.

The second part of the experiment can be worked out as a group average, to show the typical origin of the super-ego for the group. Patterns of scores for individuals can also be examined, to see what the origins are for each person. It could be discussed why some introject the demands of the group in preference to those of their parents.

The result of this part of the investigation makes most sense in terms of a theory of introjection of the demands of other people – a complex form of learning which takes place under certain conditions.

Summary. Subjects are given an indirect measure of self/super-ego discrepancy, and a measure of guilt feelings. A positive correlation is found between these, and the theoretical implications are considered. Semantic measures of the contents of the super-ego are compared with the demands of friends and of parents.

APPENDIX I

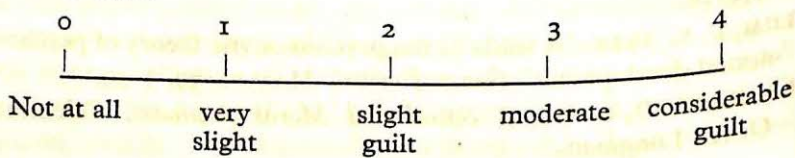
happy	:	:	:	:	:	:	:	sad
calm	:	:	:	:	:	:	:	excitable
strong	:	:	:	:	:	:	:	weak
ordinary	:	:	:	:	:	:	:	unusual
dishonest	:	:	:	:	:	:	:	honest
active	:	:	:	:	:	:	:	passive
small	:	:	:	:	:	:	:	large
striking	:	:	:	:	:	:	:	plain
cruel	:	:	:	:	:	:	:	kind
impulsive	:	:	:	:	:	:	:	placid
soft	:	:	:	:	:	:	:	hard
colourful	:	:	:	:	:	:	:	dull
dirty	:	:	:	:	:	:	:	clean
fast	:	:	:	:	:	:	:	slow
delicate	:	:	:	:	:	:	:	rugged
tense	:	:	:	:	:	:	:	relaxed
fair	:	:	:	:	:	:	:	unfair
rigid	:	:	:	:	:	:	:	flexible

APPENDIX II

SURVEY OF MORAL ATTITUDES

Indicate to what extent you feel worry or guilt about your own behaviour in the areas listed.

Use this scale.



MOTIVATION AND CONFLICT

1. Being untidy
2. Wasting money
3. Not keeping fit
4. Drinking alcohol
5. Being unkind to people
6. Smoking
7. Not telling the truth
8. Boasting
9. Not being punctual
10. Selfishness
11. Physical cowardice
12. Not going to church
13. Sexual activities
14. Not working hard enough
15. Lack of self-control
16. Causing suffering to others
17. Not keeping promises
18. Not giving money to good causes
19. Lack of persistence
20. Not keeping clean

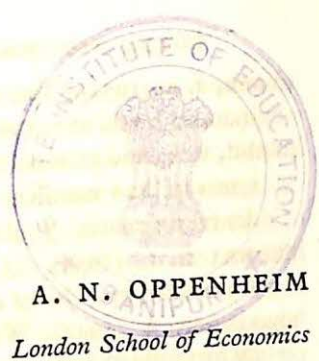
REFERENCES

- A.D.I. American Documentation Institute. Address: Chief, Photoduplication Service, Library of Congress, Auxiliary Publications Project, Washington 25, D.C. Order Document No. 5500 and reference number.
- ATKINSON, J. W. (1957). 'Motivational determinants of risk-taking behaviour'. *Psychol. Rev.*, **64**, 359-72.
- ATKINSON, J. W. (ed.). (1958). *Motives in Fantasy, Action and Society*. New York: van Nostrand.
- ATKINSON, J. W., and MCCLELLAND, D. C. (1948). 'The projective expression of needs. II. The effect of different intensities of the hunger drive on thematic apperception.' *J. exp. Psychol.*, **38**, 643-58.
- BLUM, G. S. (1949). 'A study of the psychoanalytic theory of psychosexual development'. *Genet. Psychol. Monogr.*, **39**, 3-99.
- BRAHMACHARI, S. (1954) *Social and Moral Attitudes*. Calcutta: Orient Longmans.

- CLARK, R. A. (1952). 'The projective measurement of experimentally induced levels of sexual motivation'. *J. exp. Psychol.*, **44**, 391-9.
- CLARK, R. A., and SENSIBAR, M. R. (1955). 'The relationship between symbolic and manifest projections of sexuality with some incidental correlates'. *J. abnorm. soc. Psychol.*, **50**, 327-34.
- CRONBACH, L. (1953). 'Q technique', Chap. 13 in Mowrer (ed.): *Psychotherapy. Theory and Research*. New York: Ronald.
- FESTINGER, L. (1942). 'Wish, expectation and group standards as factors influencing level of aspiration'. *J. abnorm. soc. Psychol.*, **37**, 184-200.
- FREUD, S. (1923). *The Ego and The Id*. London: Hogarth Press.
- LESSER, G. S. (1957). 'The relationship between overt and fantasy aggression as a function of maternal response to aggression'. *J. abnorm. soc. Psychol.*, **55**, 218-21.
- LEWIN, K., DEMBO, T., FESTINGER, L., and SEARS, P. S. (1944). In J. McV. Hunt (ed.): *Personality and the Behavior Disorders*, Vol. 1. New York: Ronald.
- MCCLELLAND, D. C. (1951). *Personality*. New York: Sloane.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., and LOWELL, E. G. (1953). *The Achievement Motive*. New York: Appleton-Century-Crofts.
- MCCLELLAND, D. C., CLARK, R. A., ROBY, T. B., and ATKINSON, J. W. (1949). 'The projective expression of needs. IV. The effect of need for achievement on thematic apperception'. *J. exp. Psychol.*, **39**, 242-55.
- MURRAY, H. A. (ed.) (1938). *Explorations in Personality*. New York: O.U.P.
- MURRAY, H. A. (1943). *Thematic Apperception Test*. Cambridge: Harvard Univ. Press.
- OSGOOD, C. E., SUCI, G. J., and TANNENBAUM, P. H. (1957). *The Measurement of Meaning*. University of Illinois Press.
- SANFORD, R. N. (1936). 'The effects of abstinence from food upon imaginal processes: a preliminary experiment'. *J. Psychol.*, **2**, 129-36.
- SEARS, R. R., MACCOBY, E. E., and LEVIN, H. (1957). *Patterns of Child Rearing*. New York: Row, Peterson.

Chapter 3

COMMUNICATION



A. N. OPPENHEIM

London School of Economics

The field of communication studies embraces every form of transmission which is not genetic. It includes, or should include, all forms of art, communication between animals, mime, schoolteaching, psychotherapy, the work of boards and committees, linguistics, historical research, mathematics, and a host of other things, in addition to face-to-face conversations and the influence of the media of mass communication, to which the term is most commonly applied. One might say that the capacity for communication, in its great subtlety and variety of languages, across great spans of time or distance, is the most distinguishably *human* form of behaviour.

Psychologists are not the only ones to study communication. Physicists and engineers are interested in developing new means of transmission. Information theorists and experts in linguistics study the amount and nature of precise, or imprecise, information contained in the messages which may be transmitted. Advertisers are interested in the effects which such messages produce in the respondent. The field is so vast and so complex, and systematic theoretical studies so few and far between, that almost the first thing we shall have to do is to limit ourselves in some way, and to consider how the particular aspect of communication which we propose to study fits in a more general framework.

It is often found convenient to look at the process of communication in terms of several components. Thus, we may distinguish the *source* and the *recipient* of each communication, and the *message* itself. We may look also at the flow of the communication process itself, and

describe it as one-way (such as we find in the mass media) or two-way (as in an ordinary conversation). While such analyses may be helpful, we never pretend that these components can really be anything more than somewhat artificial distinctions. Thus, the recipient of a communication should not be thought of as passive; he will interpret what he hears or sees in the light, not only of the contents of the message, but also of his knowledge of the source, and he may himself become a source of communication even before he has completely received the message. A newspaper columnist may seem to be engaged in a one-way process of communication, yet he receives mail and he has an image of his public in mind as he writes. Above all, since communication is such an intensely human activity, it takes place at several levels simultaneously, and each message is differently perceived by each recipient. Thus, we note not merely the speaker's words, but his dress and gestures also, and his message will be received by us with scepticism or enthusiasm. If therefore in each of our experiments we focus on one particular aspect of the communication flow, we shall do well to remember that these are but abstracted parts of a complex dynamic process.

In what follows we shall limit ourselves to the more superficial features of communication between humans by means of language; to verbal behaviour, in other words. One experiment will examine the efficiency of human beings as transmitters, another will help us to look more closely at the contents of the message itself, and a third will illustrate the kind of thing we do with a communication once it has reached us. For the purposes of this book, a selection has obviously had to be made; in particular, we are not going to take up the *effects* of communications on their recipients.

A STUDY OF RUMOUR

Introduction. In everyday life, human beings often receive, store, and transmit information, within varying time spans. How good are we at accepting, remembering, and reproducing verbal messages? Most of us believe ourselves to be reasonably accurate in reporting a conversation, in passing on information or in repeating an announce-

ment. Yet we are all familiar with the fact that memories are fallible, and especially so over a period of time. Moreover, something depends on how we interpreted the message, on what we thought the communicator was trying to say. And when we come to repeat what we heard, we may adjust our story to our audience, we may lend it a somewhat different emphasis or perhaps embroider a little on it. These processes go on whether the message is rumour or fact. They have an important bearing on the accuracy of testimony in a court of law.

The method we shall use in this experiment is generally referred to as 'Serial Reproduction', and it is reminiscent of a well-known party game. Sir Frederic Bartlett used it in his studies on memory processes (1932), and since then it has been used by many investigators in the field of learning and recall. Allport and Postman (1958) applied variations of this method specifically to the study of the transmission of rumours in war time. We shall not concern ourselves here with the motives for passing on a rumour, or for believing a rumour or modifying it. We are more interested in the kind of distortions and omissions that take place even under optimal conditions.

The empirical problem. It is often difficult to bring a piece of social behaviour into the laboratory for study, without distorting it. Some investigators have successfully started rumours among groups of students, and have observed the changes which the story had undergone when it was recounted to them. In such studies one usually lacks adequate controls over time intervals, choice of subjects, etc., and one misses the various intermediate stages of change. For this reason most investigators prefer to forgo 'natural' conditions, and instead use small groups of subjects who pass on a chosen story from one to the other in a laboratory setting. In this way, the various intermediate stages are observed and can be recorded, and there is control over the sequence of subjects and over the time intervals. We should keep in mind, therefore, that in such experiments we obtain a level of accuracy which is most likely far in excess of that which prevails in everyday life, since the subjects are usually of more than average intelligence and are highly primed to reproduce the contents of the message as accurately as possible. Also, the time intervals are usually very brief.

Sometimes such experiments are combined with a demonstration. A picture may be thrown on the screen before an audience and one of them, or the experimenter, describes it to the first subject, who is so positioned that he cannot see the screen. He, in turn, passes on what he has heard to a second subject, who also cannot see the screen, and so on – while the audience continues to watch the picture and to compare it with each successive account.

Aim of the investigation. Our first aim will be to see to what extent messages can be correctly reproduced by passing them on from one person to another, when all those involved are doing their utmost to produce 'total recall'.

Afterwards, we shall want to study the types of errors, distortions and omissions which take place, and to draw some conclusions from these data concerning the accuracy of communication between people, the transmission of rumour, and the accuracy of testimony.

Apparatus. Two connecting rooms and a passage, or three interconnected rooms. A tape recorder, or else a shorthand writer. Two or more passages of prose, or two pictures, or some other stimuli. Sample passages are given (p. 85) but the Instructor may wish to substitute two of his own choice.

Method. The subjects are assembled in another room, or in the passage, and allotted numbers in any order. They are told that the purpose of the experiment is to test the accuracy of their memory, that they will be told a story to which they must listen very carefully, and that they will then be asked to repeat this story immediately to the next person in order. There should be some emphasis on getting every aspect and every detail of the story just right.

The first subject is now called in. The story is read to him slowly, or the stimulus object is described to him without allowing him to see it. The second subject is then called in, and the first subject repeats to him what he has heard. This is recorded on tape, or by the shorthand writer. The first subject now leaves the room, but since he should be kept out of contact with the remaining subjects, he will be asked to leave by another door into a different room. A third subject is now called in, and the second reproduction takes place, and so on, until all the subjects are assembled in the last room.

Chains of seven to ten reproductions tend to be the most useful; after that, the story has usually become so shortened that little further distortion will take place. Larger classes could be divided into several successive chains, with different (or identical) first stimuli. If time permits, each group can run through the sequence twice or more, with different materials. With very large groups, the audience situation mentioned above can be used.

Almost any kind of prose or pictorial material can be used. Prose passages of about 150-200 words are quite effective, and the experimenter should consider beforehand what particular aspects he wishes to demonstrate. If, for instance, loss of detail, or loss of individualistic style, is to be demonstrated, then the passage of choice should contain numerous details of description, or be written in vivid style. If racial prejudice is to be demonstrated, then the material should offer opportunity for distortion to bring it in line with racist expectations. It is also often useful to have fairly ambiguous materials (e.g. an account of a road accident without stating whose fault it was) to show how readily closure and conclusion-drawing take place.

For convenience, two prose passages - extracts from a magazine - have been included at the end of this section. Both have been found useful with university students. Passage (1) offers a certain amount of detail, but rather stereotyped style, and some of the statements are often distorted to bring them into line with the subject's own political convictions. Note that the speaker's voting preference is not stated. Passage (2) contains more detail, of a descriptive kind, and it has been chosen particularly for its more vivid style. Both passages are perfectly 'fair', in that they do not contain highly abstruse or foreign material, which might be more difficult to memorize.

Treatment of results. No statistics need to be used. The class is recalled, and the successive reproductions are played back or read out to them. This in itself can prove very effective and illuminating.

The class may be led into a discussion of the various types of errors, e.g.:

- (a) condensation - the stories get shorter and shorter, often with large chunks of material left out;

- (b) fabrication – when uncertain, subjects do not hesitate to ‘fill in gaps’ in their memory, in spite of the stress on accuracy in the instructions;
- (c) inability to produce gaps – no matter how fragmentary in the memory, each story will tend to be reproduced as sets of complete sentences, and not as ‘He had just come in, eh, something-something, eh, eh, school, I think’;
- (d) loss of style – the characteristic style of the original passage will usually be lost;
- (e) loss of detail – it is important to note, not merely that detail is lost, and which details are lost, but also that what is ‘detail’ and what is ‘essential’ implies meaningful interpretation, an independent activity on the part of the subject, and in a sense the opposite to ‘photographic’ remembering;
- (f) highlighting and isolation – sometimes a detail or a name, because of its position or its oddity, will be made to stand out and be reproduced faithfully long after other details have been lost, often in a seemingly meaningless fashion;
- (g) distortion – consider especially distortions to make the story fit in with known facts, or with perceived ‘causes’, or with the subject’s own prejudices: the substitution of what the subject thought they ‘must have meant’ in place of what he actually heard;
- (h) efforts after meaning – various attempts to make the story more meaningful or more believable, in line with the subject’s own needs and stereotypes.

Other types of error may be noted, and subjects may also be asked to discuss what accounts for the fact that some parts of the story were nevertheless reproduced faithfully.

Summary and conclusions. The class should consider to what extent the entire process of remembering may be interpreted as an Effort after Meaning (see Bartlett, 1932). In the light of this, a detailed summary of the types of errors observed might be given, together with an account of need-determined distortion.

Practical applications. This experiment highlights the fact that human reception and transmission of messages is by no means a

passive process, or one that is prone to random errors or attrition. On the contrary, the process is organized, dynamic and meaningful to the individual, the types of errors and distortions can be studied, and the errors are not merely errors of omission but also of fabrication. In teaching, propaganda, advertising, etc., these facts have to be faced when attempting to produce specific results.

The implications of this experiment for the acceptance of evidence in a court of law should also be considered.

Passage 1. As I'm a schoolteacher, I'm looking for the party which suggests the most constructive way of dealing with our educational problems. Of course, both parties talk about the obvious things – smaller classes, more teachers and so on. But I believe that there is something radically wrong with secondary education, and the cause of it is the split at eleven plus. It is quite unfair to judge children on one exam at that age and the exam places a great strain on them.

I believe in comprehensive schools and I will vote for the party which is most likely to provide them. At present the clever children and the dullards are probably all right. It is the borderline cases which suffer. If you run various courses within one comprehensive unit it is much easier to transfer children from one course to another, and to make sure that the children are getting the education which suits them.

What else? Well, naturally, I am for the party which is most likely to cut prices, and to bring about disarmament. And, although the principle of equal pay is now established in our profession, I'd like it to apply everywhere. But education is my main interest.

Passage 2. Hum a melody – and Alan – only seven – can sit down at the piano and rattle it off. For this miner's son is a phenomenon. He has never had a lesson, and can't read music. But since he was three he has been banging out tunes by the score.

We found him at his grandfather's house in Chestnut Street. He had just come in from school and was bent over a colouring-book – red jersey, scrub of fair hair, grubby hands. We had with us a sheet

of music, which he glanced at with uncomprehending disinterest and then went back to scribbling. We also had with us a newly-issued record of a French folk-tune. His mother put it on the gramophone. Alan pressed on with crayoning a cow green, without paying any apparent attention to the music. It finished and his mother said, 'Go on, Alan, play it. And stop picking your nose.'

Looking pretty bored with the whole thing, Alan reluctantly and slowly got on the stool. Then it seemed that a spring had suddenly shot apart inside him. All four limbs began jumping and jabbing, head bobbing, body jerking – and the melody was coming off the keyboard like several Jaguars flat out. Nor was it that he had instantly seized the tune – but it was being repeated with all the frills and variations of the orchestration on the record.

CONTENT ANALYSIS

Introduction. Each communication, whether verbal or pictorial, in mime or gesture, has contents of some kind. We tend to think that the contents of a message have been put there by the communicator, but a good deal will also depend on who the communicator is, and how he expresses himself. Ideally, we might like to distinguish between the ideas in the communicator's mind, the actual contents of the message, the context within which the communication takes place, and the way in which the communication is received by the subject. Often this process is deeply influenced by needs and emotional variables; this is particularly so where the contents of the message are meant to be persuasive. Thus, a father who wishes to discourage his son from borrowing money may prepare a little statement on the virtues of saving; his paternal method of delivery, and the context of rebellion and discipline within which it takes place, may cause the son to regard the message as the splutterings of a spoilsport. In this process, the few selected sentences on the virtues of saving constitute only a fraction of the contents of the communication, and by studying these in isolation we should probably be mistaken in predicting what effect they would have on the subject.

Nevertheless, not unreasonably, great attention is paid in the study

of communication to the analysis of the contents of messages. We realize that this is only part of the story, but it is a starting-point, at least. We can begin by doing some sort of detailed analysis of the message itself, and then expand our concern to the various contextual variables, to the intention of the communicator, the needs and emotions of the subject, etc. We could start with some very simple methods of content analysis, and then try to develop more subtle approaches. In many situations a surprising amount can be learnt from a relatively crude and simple analysis.

Most of the methods of content analysis have been developed before the Second World War, and were used during the war in studies of propaganda. After the war, attention was focused more often on the mass media, and there is a rich variety of studies taking as their raw material the daytime radio serial, fiction in women's magazines, a bond sales drive, advice on child upbringing, Shakespeare's use of imagery, atrocity propaganda, Father Coughlin's speeches, television programmes in New York, biographies in popular magazines, German films between the wars, senators' mail, the treatment of foreign news in provincial newspapers, and personal and clinical documents. Probably the most useful text on the method of content analysis is by Berelson (1952), whose classical study of heroes and villains in magazine stories has done much to popularize the use of the method (1946). Lasswell and his colleagues (1949), and White (1949), show the application of content analysis to the studies of propaganda. Wolfenstein's (1951) work on the changing fashions in child rearing advice is particularly illuminating. In this country, the method has been used by Himmelweit, Oppenheim and Vince (1958) in their studies of television.

The empirical problem. The chief requirement of a method or system of content analysis can be summarized by saying that it should permit *objective quantification* of the raw data. The requirement of objectivity means that different analysts could use the same system on the same data and obtain substantially the same results. This type of objectivity becomes more difficult to secure as we move beyond the realm of manifest content into the areas of values, emotional force, underlying intent, etc. The more the analyst is required to interpret

the data before classifying them, the more difficult it becomes to ensure objectivity.

How to quantify content data can best be seen in terms of the choice of dimension, and of units of analysis. The *dimension* will be determined by the focus of interest of the study, e.g. the amount of violence in TV plays, or the presence of anti-American attitudes in certain newspapers, or the amount of tension reflected in a given series of personal documents. The *units of analysis* are 'bits' of content that can be identified and counted. Thus, in studying the amount of violence in TV plays, shall we count each shot and each blow and add them up, or shall we consider each fight as one unit, or shall we include verbal altercations as well, and count each term of insult or abuse? And in the analysis of newspaper policies, shall we count each word with an anti-American connotation, or each whole sentence, or each 'theme', or each paragraph, or each inch of column space? The choice of units depends on the problem under investigation, on the type of data available, and on the amount of detail that is desired.

The actual *system of classification* used will also differ from each investigation to the next, depending on the nature of the problem. Thus, newspaper contents can be classified in terms of their anti-Americanism, their glorification of sex and violence, their support for organized religion, their political leaning, etc. Such classifications may take place at different levels of interpretation and with a combination of more than one dimension, or with units of different sizes, provided always that the need for objectivity is kept in mind.

Aim of the investigation. Within the broader purpose of familiarizing the student with the methods and problems of content analysis, we shall outline three small, independent investigations:

- (a) an analysis of radio or television programmes in such broad terms as sport, drama, news, music, serious programmes, panel games, etc., with a view to comparing two or more stations or networks over a period of time – for instance, comparing BBC and ITV programmes for the same period, and perhaps comparing both of these to available data for the U.S.A.;

- (b) an analysis of the personality characteristics of heroes and villains in mass-media fiction, such as women's magazines, Westerns, etc.;
- (c) an attempted analysis of (implied) values in some selected radio or television programmes.

The description will be kept in fairly general terms, to allow the class to choose their material in accordance with what is available and what may be topical.

Apparatus. A set of *Radio Times* and *TV Times* magazines, or their equivalents, covering a given period.

A set of a particular women's magazine; or a set of scripts of a Western film series or TV presentations; or facilities for showing a set of such films in class.

A tape recorder.

Stencilled copies of content analysis instructions and sheets for recording.

Method. For the first enquiry – a comparative study of time devoted to different types of programmes by different networks of radio or television – each programme will be classified separately, and its duration (in minutes) will be recorded. Decisions will have to be made concerning the period of time to be covered, the inclusion and exclusion of certain types of programmes (e.g. children's programmes), and the system of classification to be used. With minor emendations, the system used by Smythe (1953) could probably be used, and his definitions adopted (see his Appendix B), especially if it is intended subsequently to compare the findings with American data. It will, however, not be necessary to view or listen to the programmes. It would probably be best if each programme's title and duration were recorded on a separate card, the cards classified, and the times added up. This would also allow for a check on the reliability or objectivity of the method. It is important to note that types of programmes differ with the different times of the day, and with the different days of the week, and also that many programmes run on weekly, fortnightly or monthly cycles; a period of sufficient duration should therefore be covered in the analysis. Refinements readily

suggest themselves, such as a comparison between early and late programmes, or between week-days and week-ends.

For the second study – a comparison of fiction heroes and villains – only plays or stories should be used. The material for analysis could consist of a number of issues of a women's magazine, or a set of films, or scripts of films. In each story the main characters should be identified, and rated on a number of personal characteristics. A start could be made with such simple attributes as age, sex, race or colour, cleanliness, looks, social class, marital status, accent or ethnic origin, body build, etc., each rated on a five-point scale whenever possible. Even this superficial analysis can sometimes be quite revealing. One could then go further and deal with personality ratings, such as used by Himmelweit *et al.* (1958) (Chapter 14 and p. 463) following Smythe (1953), an example of which is given at the end of this section. It will be necessary to prepare stencilled rating sheets, and to have each character rated at least twice.

For the third study – an analysis of values – it is best to arrange beforehand that all the members of the class will view or listen in to a particular programme on a particular evening at home, and also to have the programme sound-recorded on a tape recorder. When the class meets subsequently, the recording can be played back and a collective attempt at analysis can be made. This will be a more ambitious exercise, which could show the class some of the problems involved in doing this kind of analysis, the various levels of interpretation, the problem of values which are implied or taken for granted but never explicitly stated, the problem of emphasis through the technique of presentation, etc. The system developed by White (1949), or the one used by Himmelweit *et al.* (1958) (Chapter 17), could be employed and has been given at the end of this section. Again it will be necessary to prepare stencilled rating sheets.

To give a very simple example, suppose that the class has looked at some part of the Olympic games, say a long-distance athletic event. There is, of course, the factual, surface-level reporting of who won the race, what the precise time was, etc., but we may raise some further questions concerning the values implied or stated, e.g. that sport is a 'good thing' and worthy of time, money and effort; that athletes

from all over the world are like brothers; that competing for one's own country is 'a good thing'; that fair play is very important; that athletes dress in certain ways, and not in others; that men and women compete separately; that there are proper and improper ways of criticizing selectors and organizers; that patriotism is 'a good thing'; that teamwork is 'a good thing', etc. etc. We may further ask how these values are brought out: are they shown pictorially? does the commentator infuse tension (by raising his voice and speeding up his diction) into what would otherwise be a rather dull spectacle? does the camera seek out a national sports hero? does the audience convey its feelings? etc. etc.

While it could be argued that values are conveyed, to some extent, by *every* type of programme, they are probably most richly presented in dramatic performances. The main plot will have such underlying themes as 'crime does not pay' or 'pride before the fall' or 'a woman's place is in the home', but in addition we should note, say, that the characters consume alcohol in various ways; that they reflect certain attitudes towards money or animals or their parents; that they act out the roles of e.g. the rebellious adolescent daughter, the humble clerk, the butler, the tycoon, the indulgent father, the chirpy Cockney, etc.; that they seem to live in a world which is easygoing, or pitiless for the weak, or dangerous for a woman alone, or in which virtue is equated with foolishness; and so on. Plays come and plays go, but often their 'atmosphere' or 'genre' persists as a type for many years and their values, though never stated explicitly, are tacitly implied.

Statistical treatment and results. The results of the first enquiry will simply be the total time (in minutes), or the average time per day or per week, given by each network to each type of programme, as classified. Direct comparisons can be drawn, provided the same period of time has been covered, and these can be made more obvious by means of histograms. If desired, the raw data can be reworked to percentages of an average broadcasting day, or of some other base, and the results expressed in a bloc diagram. The analysis should make clear to what extent and in what categories of programmes the networks differ from one another. If possible,

comparisons with data from the U.S. may also be attempted, provided the same system of classification has been used.

The results of the second study will be a set of ratings on five-point scales, for a group of heroes, heroines and villains. As a start, simple averages could be computed for each of the three groups on each personal attribute. Sometimes it is useful to consider the spread of the ratings as well as their mean, to find out whether, say, villains are more stereotyped and uniform in certain characteristics than heroes. It can be quite rewarding to draw three overlapping little histograms (for heroes, heroines and villains, respectively) for each rating scale. The results can be used to show to what extent, for instance, villains are presented as foreign-born, swarthy, cowardly and alcoholic, or heroes as fair-haired, violent, unmarried and well-educated. Apart from a comparison of heroes and villains, the results can also be used to highlight the value system held by the particular film or magazine producers.

The results of the study of values are not always amenable to statistical treatment, though each set of values could be scored as present/absent, or as rewarded/unrewarded. The session will have achieved its goal if it has opened the students' eyes to the relative crudeness of the preceding methods of content analysis, and to the difficulties in developing a system that is both subtle and objective.

Summary and conclusions. From each of the three small studies suggested, a number of specific conclusions can be drawn, relating to the material which has been analysed. In addition, some broader questions might be considered, e.g. the degree of objectivity (reliability) which can be obtained; the problems of taking samples of data for analysis out of a much larger universe; and the extent to which categories must be built up to suit the particular problem at hand.

Practical applications. By far the widest application of methods of content analysis has been to the study of the media of mass communication. First of all we have in mind here the studies of *effects*, and the prediction of such effects. In the fields of advertising, propaganda, documentary films, etc., some definite effects are usually aimed at, and content analysis can help to show how and why such

aims are achieved, or not achieved. This involves also the 'other side of the coin' of content analysis: the subject's perception of the contents, which may be very different from the message intended.

This is also related to the study of *methods of presentation*. For instance: in educational films, how are certain points best put across, and at what speed? Or: Is it better to present one side only, or both sides of a controversial question, when trying to persuade people to a particular point of view? Or: How does the use of close-ups influence the dramatic impact of a film or television programme?

Content analysis can also tell us something about the *source* of the communication. It can help us to answer such questions as: Between which members of an alliance is the propagandist trying to sow dissension? What stereotypes of his audience does he use? What are the values implicit in his presentation? What is his image of himself, in relation to his audience?

Content analysis is also widely used as a research tool, in the study of historical or clinical documents, the deliberations of small working groups or committees, the processes of psychotherapy, and the works of well-known authors or poets.

CHARACTERISTICS OF HEROES AND VILLAINS

(Five-point rating scales)

Goodlooking	Ugly
Clean	Dirty
Tidy	Untidy
Physically Strong	Weak
Physically Healthy	Unhealthy
Physically Gentle	Violent
Courageous	Cowardly
Just	Unjust
Kind	Unkind
Sharp-witted	Dull
Enterprising	Unenterprising
Assured	Insecure
Educated	Uneducated

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

Polite	Impolite (refers to manners, courtesy)
Satisfied	Dissatisfied
Honest	Dishonest
Materially Generous	Mean
Loyal	Disloyal
Idealistic	Materialistic
Trusting	Distrusting
Modest	Boastful
Selfless	Selfish
Strong in Character	Weak in Character
Warmhearted	Callous
Industrious	Lazy
Happy	Unhappy
Humorous	Dour
Sane	Neurotic or Insane

ANALYSIS OF VALUES

<i>Value:</i>	<i>Expression:</i>
HONESTY (INTEGRITY)	'He's straight'
TRUTHFULNESS	'It never pays to lie'
JUSTICE	'He always takes an unbiased point of view'
PURITY (SEXUAL)	'She's a decent, clean living girl'
REGARD FOR LIFE	Do not kill, regard for safety of life.
UNSELFISHNESS	'He takes the refuse down for the lady upstairs twice a day'
RELIGION	'He always insisted on saying Grace before each meal'
RESPECT FOR FAMILY LIFE	Unity of family, bring friends home, keep parents informed of own whereabouts
LOVE OF HOME	'There's no place like it'
RESPECT FOR PARENTS	Admire, look up to
LOVE OF CHILDREN	'It's only their child which keeps Effie and Jim together'
MARRIAGE IS A GOOD THING	'Poor thing, she's going to be left on the shelf'

<i>Value:</i>	<i>Expression:</i>
OBEDIENCE	'You never have to tell Tommy to do anything twice'
CONVENTION	'She wouldn't be seen dead in church without a hat on'
GROUP UNITY	Country/Europe/World/Sects/ Patriotism
MODESTY	'He never talks about his achievements'
GENEROSITY (MATERIAL AND OTHER)	'He gave 10 million to charity.' 'He gave a great deal of time in the consultation of this case'
TOLERANCE (INCLUDING RACIAL)	'It's difficult for him because he always sees everybody else's point of view'
FRIENDSHIP (BEING FRIENDLY, SOCIABLE INCLUDED)	'He supported her through thick and thin'
DEPENDABILITY	'You can always rely on him not to let you down'
DETERMINATION	'She will never rest until she's won the award'
ACHIEVEMENT	'That is a man who has got on in life'
INDEPENDENCE (NATIONAL AND PERSONAL)	'It is impossible for a man who must go his own way to live in this oppressed country'
SECURITY	'I would not risk giving up my job without another one to go to'
DOMINANCE	'They flourished under the firm discipline of their leader'
INTELLIGENCE	'It is a pity that he is not as bright as his brother'
KNOWLEDGE	'If you want to get the best opinion, get an expert in'
WORK (INCLUDING SKILL, CAPABILITY)	'He's very good at his job and would not dream of missing a day'
CREATIVE	'She has a most vivid imagination and has designed a backcloth'
OWNERSHIP	Possession of object(s), wealth. 'Apart from his estates he had the finest collection of jade in the country'
MONEY (PLUS ECONOMIC)	'He worked overtime in order to buy a refrigerator'
MANNERS	'Say thank you, Tommy'

Value:

Expression:

CLEANLINESS

BEAUTY

APPEARANCE

'Her home shone like a new pin'
Of people, objects, nature, literature,
music, etc.
Goodlooking, well turned out, taking
care of self, etc.

STEREOTYPES

Introduction and literature. Any communication is usually based on some form of understanding between the source and the recipient. There must be some form of common language, in the wider sense of the word, and this may be used to persuade, to transmit information, or to produce an emotional state in the subject. In this process the recipient of the communication should not be thought of as passive; indeed, it can often be said that he meets the communication half-way, and this is especially so in the case of the mass media. The speaker, writer or player can rely on a common cultural frame of reference, and can make use of many preconceived images and ideas in order rapidly to sketch a situation or portray a character. We are familiar, for instance, with the use of stock characters in plays, jokes, and cartoons. On the other hand, a European audience viewing an Indian or Japanese film may fail to get the full flavour of the communication because of the absence of these frames of reference.

Walter Lippmann (1922) was the first to call these 'pictures in the mind' *stereotypes*, and the term has gained wide currency. Usually, stereotypes are regarded as a 'shorthand in thinking', a set of role expectations which is fairly fixed and widely held, and which may determine selective perception. Research in this field has most often been concerned with racial and national stereotypes (e.g. Katz and Braly, 1933) and their relation to prejudice (e.g. Blake and Dennis, 1943), but we meet with stereotypes in most other spheres of life. We have, for instance, stereotypes about the different social classes, stereotypes about different occupational groups (such as plumbers or lawyers), stereotypes about family members (such as mothers-in-law), about historical figures, about criminals, about people from

other parts of the country, about far-away places (such as the Pacific islands, Hollywood or Paris), about labour and management, about the elderly, and about ourselves.

Stereotypes can often be dangerous if they are believed too literally. Stereotypes are usually false, and can often be shown to be false quite easily, yet they find their way into our thinking and attitudes quite insidiously. In particular, they usually lead us to make generalizations which are too sweeping, and blind us to the many exceptions and differences which exist among the members of a group. Stereotypes usually reflect an attitude, either friendly or hostile, and they often take the form of convenient rationalizations for prejudice against others, and for believing oneself or one's own group to be superior. Many stereotypes have become part of the common culture and are adhered to, less because of any deepseated prejudice than for the sake of conformity; they are then often perpetuated and given wider currency by the mass media. In this way a vicious circle is sometimes set up.

Most of us believe that we do not use stereotypes in our thinking, and that our views are generally based on experience and sound evidence. It comes as a shock to many of us when we find out how much of our thinking is based on hearsay, unconscious rationalization and baseless over-generalizations. When we travel, for instance, or when we meet people from another culture, we may find that our stereotypes led us to expectations and preconceived ideas which were wrong. Yet often we do not discard these stereotypes but continue to adhere to them, making 'an exception which proves the rule' out of our experience. Or else we develop a new stereotype, a false and sweeping generalization based on very limited evidence. Some of us, because we are more rigid and more intolerant of ambiguities, have great *need* to lean on stereotypes in our thinking and attitudes (Eysenck and Crown, 1948).

The empirical problem. We shall set ourselves the task of finding out what images are aroused in the subject's mind when he sees or hears such labels as 'negroes' or 'Turks'. In some cases no stereotypes will be aroused, in other cases variable and diffuse images will emerge, and sometimes highly uniform responses will be given. We

shall wish to know what attitude these stereotypes reflect, and how they are learnt.

It has been suggested by Gilbert (1951) that some popular stereotypes change over the years, and this could be investigated also.

Aim of the investigation. The aim will be to demonstrate the following points:

- (a) that stereotypes can readily be evoked;
- (b) that the stereotypes concerning some stimuli are more clearly defined than those concerning others;
- (c) that stereotypes reflect attitudes of acceptance or rejection;
- (d) that stereotypes are *not* based on sound evidence but are gross over-generalizations;
- (e) that some people more readily think in terms of stereotypes, than others;
- (f) that the contents of some popular stereotypes may change in the course of time.

Collectively these points will show how much the recipient of a communication 'brings to' the situation in the way of preconceptions and attitudes.

Apparatus or tests needed. Two lists, as given at the end of this section.

Both lists consist of sets of stimulus words, to be matched to adjectives, e.g. first names such as Charles or Mandy could suggest a person who is musical or pompous, practical, talkative, etc., or people may think of the Germans as hardworking or the British as conceited.

Method. Administer both lists to a group of subjects.

Statistical treatment and results. We shall relate the results to each of the aims set out above.

(a) '*That stereotypes can readily be evoked*'. It is expected that a considerable proportion of the subjects will have found it possible to carry out the instructions without undue difficulty. This suggests that they are aware of the existence of such stereotypes, though they may or may not actually believe in them.

(b) '*That some stereotypes are more clearly defined than others*'. For

each stimulus on the lists, compute the percentage of subjects who chose each adjective. It will be seen that in some cases one adjective receives a very high percentage of choices, indicating high agreement or unanimity, while in the case of other stimulus words the choices are more evenly divided among the adjectives, indicating a mixed or indefinite stereotype, or the absence of any general stereotype. Compare each percentage with chance expectation.

(c) *'That stereotypes reflect positive or negative attitudes'*. The drafters of List B (Unesco) designated four of the adjectives – conceited, cruel, backward, and domineering – as negative, and eight – hardworking, intelligent, practical, generous, brave, self-controlled, progressive, and peace-loving – as positive in connotation. Compute for each nation on List B the average number of positive and negative adjectives chosen. The differences will be pretty obvious, but if necessary a positive/negative ratio of 'favourableness' could be computed. Is it true that we tend to regard our own nation in the most favourable light?

(d) *'That stereotypes are not based on sound evidence but are gross over-generalizations'*. Arguments are sometimes presented suggesting a 'kernel of truth' hypothesis, according to which there might be at least some evidence to support the stereotype, perhaps in less general form. This argument could, however, not possibly be applied to List A, unless we are prepared to believe that parents know in advance that their child will grow up to be pompous, talkative, or practical; or unless we wish to assume that children grow up to resemble their first-name stereotypes. Such stereotypes are palpably false, yet they exist and are perpetuated. Can you suggest how they come about?

(e) *'That some people think more readily in terms of stereotypes, than others'*. Note that on List B each subject is allowed to refuse to answer in terms of stereotypes, and can say instead 'Impossible to characterize'. Compute for each subject the number of times he has given this response. It will now be seen that some people are clearly more willing to think and respond in stereotyped terms than others. Can you say why this might be the case?

(f) *'That stereotypes may change over time'*. Compare the results

from this experiment with those collected in 1948 (Unesco). In making these comparisons, keep in mind any differences between the present sample and the one used in 1948. See also Gilbert (1957).

Summary and conclusions. The relative crudeness of this demonstration will raise many points about method. Can this simple set of adjectives really reflect the 'pictures in the mind' which people have about one another? Is the fact that people can and will answer such questions evidence that they *think* in this way, or are the answers the results of the limited nature of the task? How many of our subjects have used the 'Impossible to characterize' category?

Could we have communication without stereotypes? In particular: could the mass media get on without them? Are stereotypes always '... false maps of non-existent territories' or are they sometimes useful, and partly true? When do stereotypes become dangerous?

Stereotypes are not inborn, they are learnt. But once learnt and established, can they change? Why do generally held stereotypes change, and how? Will individuals abandon their stereotypes in the light of scientific evidence, or will many still cling to their original ideas? Are some people more resistant to change than others? And are they perhaps also more prone to 'pick up' stereotypes?

Finally, how can we draw the line between a stereotype and various other beliefs and generalizations?

Practical applications. The study of stereotypes helps us to improve our understanding of racial prejudice, and of hostilities between nations. No study of the causes of war could be complete without paying some attention to the role which stereotypes play in the development and maintenance of international tensions.

In the field of propaganda the role played by stereotypes is now widely appreciated. Research is chiefly directed towards ways of originating or changing stereotypes, through the use of the mass media.

COMMUNICATION

LIST A

Below is a list of eight masculine names. Please choose, for each name in turn, an adjective that best describes a person of that name. Choose the adjectives out of the following list:

Kind, practical, pompous, intelligent, serious, artistic, stupid, good-looking.

Write your choice after each name.

NAME	ADJECTIVE
Charles	_____
Mark	_____
Norman	_____
Anthony	_____
David	_____
Richard	_____
Herman	_____
Adrian	_____

Now do the same for a list of eight feminine names. Choose from the following adjectives:

Practical, talkative, musical, middle-aged, fat, warmhearted, flighty, dull.

NAME	ADJECTIVE
Grace	_____
Winifred	_____
Edith	_____
Maisie	_____
Agatha	_____
Ruth	_____
Mandy	_____
Cathleen	_____

LIST B

From the list of adjectives that follows, which seem to you to describe the American people best? Select as many adjectives as you wish, and put a tick opposite each one, in the column headed 'Americans'. If you have no particular feelings one way or the other, just put a tick at the bottom of the column.

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

Now please go over the list again and select the words which best describe the French people. Put your ticks in the column headed 'French'.

Now please continue with the other nationalities.

<i>Adjectives</i>	AMERICANS	FRENCH	GERMANS	RUSSIANS	BRITISH	ITALIANS
Hardworking						
Intelligent						
Practical						
Conceited						
Generous						
Cruel						
Backward						
Brave						
Self-controlled						
Domineering						
Progressive						
Peace-loving						
Impossible to characterize						

REFERENCES

- ALLPORT, G. W. (1954). *The Nature of Prejudice*. Boston, Mass.: Beacon Press.
- ALLPORT, G. W., and POSTMAN, L. J. (1958). 'The basic psychology of rumour'. In Maccoby, E., Newcomb, Th., and Hartley, E. L.: *Readings in Social Psychology*. New York: Holt.
- BARTLETT, F. C. (1932). *Remembering*. Cambridge: Cambridge Univ. Press.
- BERELSON, B. (1952). *Content Analysis in Communication Research*. Free Press, Glencoe, Illinois.
- BERELSON, B., and SALTER, P. (1946). 'Majority and minority Americans: An analysis of magazine fiction'. *Publ. Opin. Quart.*, 10, 168-90.
- BLAKE, R., and DENNIS, W. (1943). 'The development of stereo-types concerning the Negro'. *J. abnorm. soc. Psychol.*, 38, 525-31.
- EYSENCK, H. L., and CROWN, S. (1948). 'A study of racial stereo-types'. *Int. J. of Opinion & Attitude Res.*, 2, 26-49.
- GILBERT, G. M. (1951). 'Stereotype persistence and change among college students'. *J. abnorm. soc. Psychol.*, 46, 245-54.
- HIMMELWEIT, H. T., OPPENHEIM, A. N., and VINCE, P. (1958). *Television and the Child*. London: Oxford Univ. Press.
- JAMES, H. E. O., and TENEN, C. (1953). *The Teacher was Black*. Heinemann.
- KATZ, D., and BRALY, K. (1933). 'Racial stereotypes of 100 college students'. *J. abnorm. soc. Psychol.*, 28, 280-90; also in Newcomb and Hartley, *Readings in Social Psychology* (various editions).
- KLINEBERG, O. (1950). *Tensions affecting International Understanding*. Soc. Sc. Res. Council, New York.
- LASSWELL, H. D., LEITES, N., and associates (1949). *Language of Politics*. New York: Stewart.
- LIPPMANN, W. (1922). *Public Opinion*. New York: Harcourt.
- SMYTHE, D. W. (1953). *Three Years of New York Television 1951-1953*. Urbana, Illinois: Gregory Hall.

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

- STAGNER, R. (1950). 'Stereotypes of workers and executives among college men'. *J. abnorm. soc. Psychol.*, **45**, 743-8.
- UNESCO. *International Social Science Bulletin*, 1951, **3**, no. 3.
- WHITE, R. K. (1949). 'Hitler, Roosevelt and the nature of war Propaganda'. *J. abnorm. soc. Psychol.*, **44**, 157-74.
- WHITE, R. K. (1951). *Value Analysis: The Nature and Use of the Method*. Society for the Psychological Study of Social Issues.
- WOLFENSTEIN, MARTHA (1951). 'The emergence of fun morality'. *The Journal of social Issues*, **7**, 15-25.

Chapter 4

SMALL SOCIAL GROUPS

THELMA VENESS

Birkbeck College

The title of this chapter begs many questions. How small is a small group? for example, accompanied by images of the seven dwarfs at a colloquium. Rather more profound is the question of what constitutes a 'group' as distinct from an assemblage, an aggregate, a crowd, or whatever other collective noun comes to mind. And if attention then focuses upon the word 'social', there arise fundamental questions underlying all the social sciences, about the springs of the social life of man, the source of all social behaviour, social institutions and collective life.

While it may be rash to generalize from, say, a group of children swapping marbles to international systems of trade – or vice versa, for that matter – there are obvious advantages in studying social life at the point where something like the unities of classical drama operate – where time, space and plot are manageable. The social psychologist, in making such a choice of theatre, may remind one of the man who was looking for a lost sixpence, not where he dropped it, in the middle of the road, but under the lamp-post – on the grounds that there was more light there. But the point need not be too well taken. Any science concerned with men in society must at times take account of 'a man's-eye view'; studying, that is, not the Bank of England as an economic instrument but, say, the function of 'the Club', 'the divi' and the weekly pocket money in family life. It is not by virtue of the more institutionalized aspects of his social life that man is called a social animal; these he shares with sheep no more than with hedgehogs. It is, rather, to the fact that the

average human being spends a considerable part of his working and recreational life in quite close contact with other human beings that we are referring. Some of his groups cannot easily be escaped, such as the family if he is a two-year-old, or the gang if he is a scaffolder. But many seem entirely voluntary. Is this grouping the result of blind instinct or is it a matter of habit? What does this form of social life bestow upon the individual? What are its rewards? Argyle (1957) has found it convenient to make a basic assumption that human beings like to be liked, and this aspiration, as Homans (1951) has hypothesized, can be satisfied through the familiarity coming from continuous interaction. Both these ideas about human social nature came from considering the small group in face-to-face interaction.

Here we return to our second question. It is by reference to face-to-face interaction, possibly, that a 'group' has been most satisfactorily defined. So Sprott (1958) regards the matter, for example. Such a definition has the advantage of being behaviouristic, in contrast with the many which call upon psychological qualities, such as interdependence, the pursuit of common goals, shared identification or consciousness of being a group.

The question, how many people constitute a small group? can then best be answered by considering the occasions in ordinary life on which people interact in this way. In a school of 2,000 children, we are told, not even the teachers can know everybody; certainly children do not all interact freely in groups larger than about thirty. Groups studied experimentally, that is by manipulation for control of one or two variables at a time, have usually fewer than ten people. It is with experimental groups that we shall be concerned here, but we must first acknowledge that the main source of interest in, and hypotheses about, social groups has come from accounts of real-life groups *in situ*, such as Whyte (1943) on street-corner gangs, from participant observation, Piaget (1932) and Susan Isaacs (1933) on children's groups, and Bion (1948-51) and Foulkes and Antony (1957) on group psycho-analytic sessions. Some of these examples, and work on committees, discussion groups, and conferences, illustrate the closeness of small group studies to current problems and preoccupations. Indeed, one might find a little worrying this general

burgeoning of interest in small groups, at the present time, when in certain places the way to be thought a 'regular fellow' is to make oneself virtually indistinguishable from everybody else. There is something salutary in the question that Klein (1956) suggests should guide much of our research – 'when is it worth-while to form a group?' It could be unfortunate if the impression were given that groups are studied chiefly in order so to increase man's efficiency at forming groups that eventually no one need ever be alone.

The 'natural historians' of small group studies have the advantage of dealing with groups that have a history and a future, usually related to their purpose, which does not have to be invented by the investigator. They have the disadvantage that, since the groups do have this independent and continuous existence, the method of study is largely observational and, to some degree, often impressionistic. Thus the man who is observing groups in order to get ideas about group functioning has to beware of his own distorting lenses rather more than the experimentalist who has built in procedures that permit him to show results independent of his judgements. The latter, however, in order to achieve precision, may have to sacrifice breadth, which means not only that to cover the same ground he must work harder, but also that it is a more laborious business to test the interaction of basic variables. Probably the most common criticism of experimental small group studies is that they are artificial (though artificiality becomes a thing of wonder and admiration when it pertains to a kidney or an iron lung). The criticism is directed at such features as the need to invent a task for the group, sometimes to give a fictitious reason for its existence, and most of all to the short life of many such groups, especially in experiments which demand, as Klein (1956) puts it, that both the culture of the society and the personality of the individual be disregarded, and which treat the individuals, carefully placed round a laboratory table or established in armchairs behind an observation screen, as if they had no group life beyond the present one. This is a price that most of science, certainly of psychology, has to pay for firm knowledge.

Early theorists about group behaviour, such men as Simmel, Cooley, G. H. Mead, Le Bon, Freud, were abstracting from real-life

small groups, often from that universally flourishing yet perhaps somewhat atypical primary group, the family. Their contribution, as well as examples of the experimental work they inspired, is represented in the anthology of small group studies of Hare, Borgatta and Bales (1955). Homans (1951) is of their tradition. Another well-known collection of papers is to a greater extent confined to experimental studies, that edited by Cartwright and Zander (1953), where one is more conscious, perhaps, of another pathfinder, Kurt Lewin, whose theoretical contribution was of the cook-book variety (at this point draw carefully a thick boundary line to prevent the life-space from curdling) but whose experimental ingenuity was enormous. Moreno has furnished small group experimentalists with convenient tools: sociometry is well reviewed by Lindzey and Borgatta (1954).

Eventually, both approaches must contribute to the same generalizations, and Argyle (1957) has claimed that generalizations are already emerging; about the causes and effects of group cohesion, effectiveness of competition, limitations to the exercise of leadership, optimum size for group problem-solving, by-products of communication arrangements, growth and maintenance of group standards, to mention some of the problems fruitfully studied. Nor should the two approaches be regarded as mutually exclusive: certainly the most careful experimentation can usually profit from intelligent observation.

Both techniques can be adopted for demonstration of small group studies. A one-way observation screen is an advantage to observational studies, especially for watching children, but is not essential. Of greater importance is that the observation should be guided, possibly by tentative hypotheses; certainly by a system of noting interactions, such as those devised by Bales, Chapple, Steinzor and others, reviewed by Heyns and Lippitt (1954). Pure experimental demonstrations, such as those described here, are invariably limited. A full-scale small group experiment usually involves a large number of subjects, because of individual variation and the requirements of experimental control, who are put into groups that may continue for several weeks. Where large numbers of subjects are available at the

same time, considerable space is demanded and a large number of experimental assistants. The main assumptions made here are four. (1) that the demonstrator is prepared to tolerate a certain degree of compromise and will accept the testing of limited hypotheses with a far from ideal design; for instance, using the same subjects under different experimental conditions, with all its possibility of instructive discussion about transfer effects. (2) that there is available in one place a willing subject-population of about twenty to thirty people. (3) that there are supernumeraries quick-witted and reliable enough to be briefed at short notice to act as experimental assistants to take charge of groups or assist with recording, permitting further instructive discussion about the wickedness of half-trained experimenters and the consequent lack of standardized procedure. Of course, where assistance can be enlisted in advance, this is a great advantage. (4) that the instructor is prepared to do a fair amount of preparation in advance in order that the experiment can be conducted within two hours.

No experiment described here involves elaborate devices for hoodwinking the subjects. They all illustrate influences on group problem-solving behaviour; namely, competition and co-operation, communication paths and group size. Experimental work on group problem-solving has been reviewed by Kelley and Thibaut (1954).

CO-OPERATION AND COMPETITION

Introduction. Occasionally, the attempt to explain familiar social behaviour, such as the mass phenomena reviewed by Brown (1954), throws up an idea which can be tested experimentally in small groups. Panic behaviour, as when a crowd dashes to the theatre exit on the shout of 'Fire' and jams it so thoroughly that no one escapes, is a case in point. Several theorists have been content to say that such behaviour is due to emotional contagion. Mintz (1951, also Maccoby, Newcomb and Hartley (1958)) has sought for an explanation which accords better with modern theoretical conceptions. Briefly, he concluded that the decisive factor concerns the system of rewards and punishment. Where the reward and penalty system is

known and stable, no panic need result. Where, however, one individual's self-seeking behaviour can disrupt the group reward and penalty system, the outcome will be 'non-adaptive' group behaviour, where the individual competes with his fellows instead of working towards the group goal. Emotional excitement may be engendered, but this is secondary. This is the hypothesis which, with various corollaries, he tested in a series of experiments, and which was to some extent supported. It can be regarded as a demonstration of conditions engendering competitiveness, which, in some circumstances, can be disastrous. There are other situations where competition improves performance, of course. This experiment is not a straightforward comparison of the merits of competition and co-operation, many of which can be found in the literature on group problem-solving and which cannot be summarized by a simple vote for one or the other. The present experiment involves a task which can *only* be performed successfully with co-operation.

The empirical problem. A task is needed, then, which a group can solve readily with co-operation, but in which 'traffic jams' will result if co-operation breaks down. Mintz had his subjects pull wedges out of a bottle. They could all be drawn out only if people waited their turn. A competitive attitude can be induced by rewards and fines, administered on an individual basis. A closer approximation to real-life mass panic is achieved by prohibiting subjects from talking to one another, thus preventing the adoption of a strategy. And the influence of emotional excitement can be tested by using accomplices as a rabble, to stir up the players. These are some of the variations tried by Mintz, who used a large number of groups for each condition. The main modification suggested here is to use three groups, each working under one of the above conditions and then as a co-operative team.

Aim of the investigation. Our main aim is to study a group extracting wedges from the bottle, competitively and co-operatively. By having every group do this operation several times under each condition, statistical comparison becomes possible. Our prediction, then, is that a co-operative attitude will give better performance. In addition, the group that is handicapped by being silenced when in

SMALL SOCIAL GROUPS

competition, should show an even greater inferiority relative to its co-operative performance. And finally, if Mintz is right, the third group, which is being excited by its audience during the competition

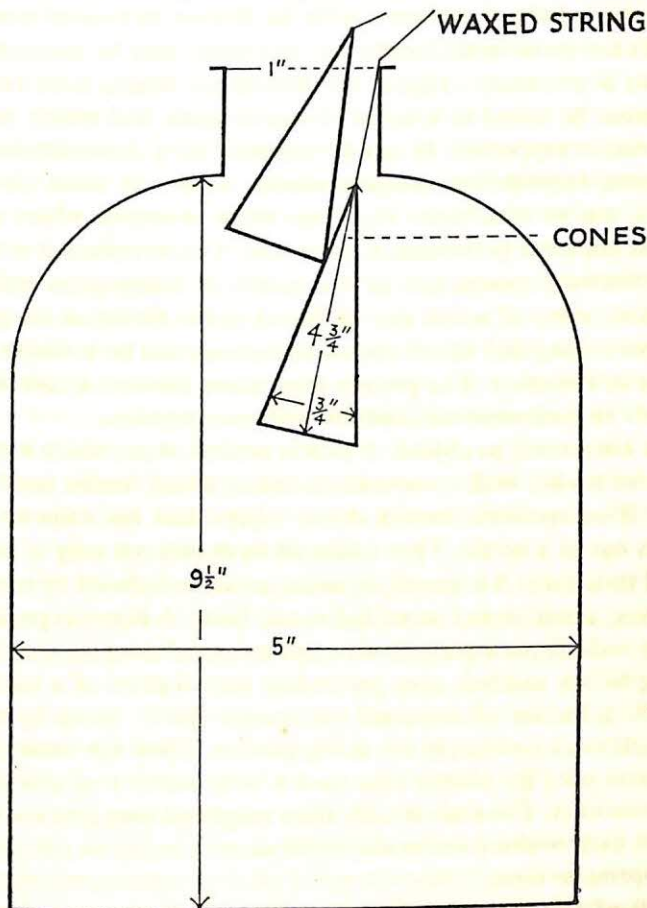


FIG. 2.

trials, should show a similar relative difference to the first group. Thus each group is being compared on its performance scores only with itself, in the absence of a large number of groups.

Apparatus. Mintz gives a diagram of the bottles and cones he

used. With the present design, three bottles only are needed: similar bottles are readily obtained from chemical apparatus suppliers. Mintz had one ingenious experimental variation which involved running water into the bottom of the bottle, a subject being fined if part of his cone were wet. A simpler system of fining is used here, so that a bottle with a spout is not required. The sponge rubber pad he had cemented to the bottom of the bottle can also be dispensed with. The bottle needs to be very securely tied down to the table legs, as even a small group in competition can muster considerable physical force. The cones can be made quite easily, of a size suitable for the bottles available and the number of subjects to be used. Mintz preferred aluminium ones, having found wooden cones apt to stick in the neck of the bottle, but an experimental assistant briefed to push them down quickly is an effective compromise. There should be two cones per subject, with spares available in case the string breaks. If each subject has two cones, then, with about six subjects in a group, the 'traffic situation' becomes closer to that of Mintz, who used 15-21 subjects. Mintz recommends the use of waxed string to prevent tangles. Each group requires a stopwatch and a score sheet. A group can be managed by one experimenter, but two are preferable.

Method. After an initial explanation to everybody that small groups are going to be formed for an experiment on group problem-solving, individuals are assigned to four groups of six, say, by random number tables. It is advisable not to have sets of people already used to working together. They must then be prevented from getting into their groups, for example by being given something to write individually, while the supernumeraries are being briefed as experimental assistants. The briefing can be done quite quickly, especially if written instructions have been prepared.

The next task is to remove the group of six who are the accomplices, and prepare them to act as a noisy audience to one working group. They are told that their job is to make a noise, exhort, sneer, in every way to try to excite - but to take care not to give advice, good or bad. And they must be serious about the job. This role is in fact difficult because it is useless if half-hearted: time to plan, even rehearse, tactics is an advantage, and their number can be enlarged with

profit. When their victim-group is working under the co-operative condition, they can sit back and watch, but silently.

The working groups are then taken to the apparatus. Two groups in a fair-sized room do not interfere with each other once they are absorbed in their problem, but the 'panic' group should be segregated. Each group will work first in competition, as the carry-over from co-operation would be too great if they started that way. Each group has about ten trials, of about twenty seconds each, under each condition. This allows for statistical comparison but is not too great a number for interest to be maintained or for learning to be perfect. These difficulties are likely to arise if every group works under *all* experimental conditions.

The three competitive conditions, one for each group, are (i) rewards and fines only, (ii) rewards and fines plus silence, (iii) rewards and fines plus the rabble. The co-operative situation which follows is the same for all groups. All groups will have the same instructions for the first part, except that the second group will be told, before and after the instructions, that they are not allowed to talk to each other or communicate in any way. Suggested instructions are:

This is a task on which you can win or lose small sums of money. You will put all your cones in the bottle and then, starting from when I say go, see if you can pull them out in 20 seconds. If you get a cone out in 10 seconds you will be credited with 3d. and for any cone out in the 20 seconds, one penny. For any cone left in at the end of the 20 seconds you have to pay a fine of one penny. At the end of each try, your wins and losses will be recorded. We shall do this several times. Is that clear? Let's get all the cones in now - touching the bottom of the jar.

The cones have to be put in one at a time, which gives a clue to the solution of the problem and which affords time to the first and third groups to plan as a group if they wish. A string is held in each hand - some subjects will soon discover that they can increase their speed by shortening the string. At the word go, a stopwatch is started and an assistant stands ready to push down the cones if they get stuck. The timekeeper calls out at half-time to facilitate the recording of

rewards. This procedure is repeated for the ten trials, rewards and fines being entered after each. It should be noted that the hypothetical monetary rewards and fines are treated with scepticism by some groups; justifiably usually. Marking rank order of pulling out the cones might be a better spur to competition for some subjects but would make scoring more laborious. The time allowed for each trial must be adjusted to fit the number of cones, and size of the bottle.

For the ten co-operative trials, the following instructions are given:

Now we are going to do this a different way. This time, you are to try to co-operate fully to get out as many as you can. There are no rewards and penalties and what will be recorded each time is the total number of cones extracted by the whole group. In this you will be compared with the other two groups working elsewhere at the same task. Again, you will have several trials of twenty seconds each. Now put all the cones in, touching the bottom, and wait till I say go.

In addition, group two is now freed from the ban on talking. In these instructions, competition is still encouraged, it will be noted, but now on a group basis; as individuals, they are encouraged to co-operate. The number of cones pulled out is recorded, and the time taken if all are pulled out before twenty seconds have elapsed.

Statistical treatment and results. Learning curves can be plotted for each group in each condition, of the number of cones removed in the twenty seconds (or of the time taken to get all out, in the second condition). The scores of each group under the two conditions can be compared by the Mann-Whitney test; and for all groups put together. By comparing the two sets of scores for each group trial by trial, a percentage improvement can be calculated, which would permit comparison of the three groups.

Summary and conclusions. Many statements of conclusions in small group experiments will come from the discussion period after the experiment, a *sine qua non* for which at least half an hour is desirable. Here subjects and experimenters will pool their observations of group dynamics, and their comments on the shortcomings of the

SMALL SOCIAL GROUPS

experimental design. An important point to have discussed is the relevance of the experimental situation to the real-life phenomenon it is meant to resemble, and what variables have been omitted in the interests of simplification. With careful recording during the experiment, it should be possible to summarize quantitative results immediately and, at least tentatively, come to conclusions about the hypotheses being tested; whether, in this situation, individually based fines and rewards create competition that impedes; whether the ban on communication increases this effect (and this will certainly be shown); and whether the arousal of excitement also contributes, contrary to hypothesis.

Practical applications. Mintz speaks of the many social situations to which the experiment has relevance. Emphasizing his point that the emotional aspect is epiphenomenal, he quotes cases where there was no panic in the complete absence of hope of escape; no incentive for anti-social behaviour. One lacks the positive application; there should be more to this than the conclusion that despair is more comfortable than struggle. Perhaps it is that in such emergencies, there is a need for clear instructions about what to do and prevention by force, if necessary, of self-seeking. The other practical application of the experiment is that generalizations about competition and co-operation must have regard to the nature of the task. Individual competition is only effective for all if it does not imply mutual interference. In a bonus payment system, for example, individual competition against an external standard is very different from the situation of one man's gain being another man's loss.

COMMUNICATION PATHS

Introduction. It has been said that an essential property of a group is that the members interact. Often limitations are imposed upon the possibilities of interaction, however. These limitations can be formal or informal. The hierarchies of a business organization, for example, determine the communication paths available for any individual. Rather similar communication networks can develop informally. According to Whyte (1943), the members of his street corner gang

tended to address all their remarks to the gang leader, and information from him was disseminated to the rest of the gang via his lieutenants. In fact, in this informal group, it was virtually the communication channels that defined the leadership of the group.

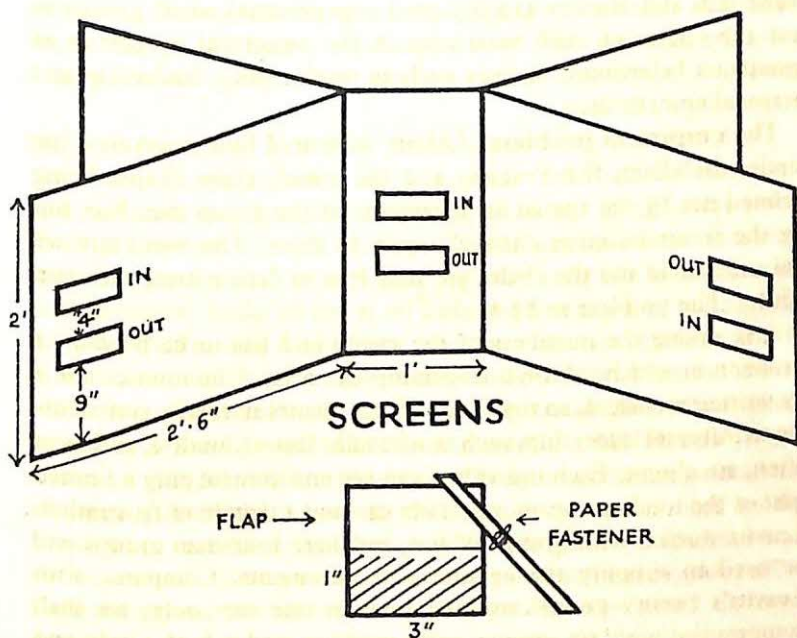
Stimulated by the theoretical formulations of Bavelas (1950, also Cartwright and Zander (1953)), Leavitt (1951, and also Maccoby, Newcomb and Hartley (1958)) used experimental small groups to test the effects of such variations in the structural properties of groups on behavioural indices such as productivity, leadership and personal enjoyment.

The empirical problem. Leavitt compared four structures, the circle, the chain, the Y shape and the wheel, these shapes being defined not by the spatial arrangements of the group members but by the communication channels open to them. The two extremes only, the circle and the chain, are sufficient to demonstrate the main effects. The problem to be worked on is one in which information is shared among the members of the group and has to be pooled – a situation in which informal leadership can arise. Communication is by written messages, so that some of the factors normally contributing to informal leadership, such as who talks fastest, loudest, and most often, are absent. Each individual can see and control only a limited part of the total operation, probably causing a degree of frustration. Leavitt worked with groups of five, but here four-man groups will be used to simplify the apparatus requirements. Compared with Leavitt's twenty groups working each in one way only, we shall compromise with six groups, each working under both circle and wheel conditions.

Aim of the investigation. Specifically, the questions to be answered by this comparison of the circle and the wheel communication patterns are three. Which is the more efficient at the task allotted to it? is the first, efficiency being measured by time taken. Ancillary questions here are concerned with the way in which a group achieves its efficiency, the content of the messages giving clues to this, and the speed with which an effective technique is adopted over a number of trials. Individual evaluations of efficiency are also relevant. Secondly, is designated leadership (from questionnaire responses) associated

SMALL SOCIAL GROUPS

with 'centrality', to use Bavelas' term – defined in terms of the smallest number of communication links needed to reach every other person in the team? In the circle design, there is no central person; in the wheel, one person is in direct communication with all others. Thirdly, how do individuals enjoy their various roles? How does the total satisfaction index (from questionnaire ratings) compare in the



DETAIL OF OPENING

FIG. 3.

two types of group, and how do peripheral and central members compare in the wheel design? And associated with this, to what extent do the participants come to understand the structure imposed on the group?

Apparatus. Leavitt, working with one group only at a time, used a 'postbox' apparatus in the centre of the table allowing messages to be sent by each individual to any other. With six groups working at a

time, simplification is needed, and this can be met by having groups of four people, who can then be separated by cardboard erections standing on the table, high enough to prevent overlooking, as in Fig. 3, with a cross-bar for communication, for the wheel design, between two people facing each other. In each of the five partitions, two small openings are cut, with flaps that can quickly be fastened back. The slots are marked In and Out to prevent collision, and they must be in reach, and large enough to permit the speedy transmission of messages, but not so large that anyone can see what is happening next door.

Each compartment bears a colour, indicated by the colour of the cards stacked ready for messages. If these are numbered in advance, some retracing of the order of events becomes possible. For convenience, let us assume that Green is facing the cross-bar and will be the hub of the wheel design, that Pink is facing him, with Blue and Yellow to right and left.

The problem material must be stacked in advance in each compartment, turned face down and numbered so that the correct one is taken each trial. Many problems are possible; Heise and Miller (1951, also in Hare, Borgatta and Bales (1955)) used a variety of interesting ones for groups of three people communicating by microphone from adjoining rooms. Leavitt used cards bearing symbols, one symbol only being shared by everybody, which had to be found. This problem is suitable for the present arrangement, four out of five symbols being on each card and the experimenter having the right answers. A stopwatch is required and a score sheet, and envelopes in which to put the messages for *each trial separately*, to permit subsequent analysis. At least two experimenters for each group are desirable.

Using one table only, Leavitt was able to introduce mechanical means of recording answers; each subject threw a switch which was registered on a master board in an observation room. This can be dispensed with by having subjects raise their hands and show the experimenter the answer written on a card. This is clearly not so satisfactory as Leavitt's method; it demands that the experimenter be on his toes lest he cause additional frustration.

SMALL SOCIAL GROUPS

At the end of the trials for each experimental condition, copies of a questionnaire are distributed to every subject. This, modified from Leavitt, first asks number of group and writer's colour for identification purposes, and then:

- 3 Did your group have a leader? If so, who was it? (answer by his or her colour).
- 4 Describe briefly the organization of your group.
- 5 How did you like your job in the group? Give an assessment on a rating scale from 1-6 with 6 indicating the greatest degree of liking. Put a ring round one of the following: 6 5 4 3 2 1.
- 6 Was there anything, at any time, that kept your group from performing at its best? If so, what?
- 7 Do you think your group could have improved on its efficiency? If so, how?
- 8 How efficient would you estimate your group to have been? 6 indicates the highest degree of efficiency: ring the appropriate number: 6 5 4 3 2 1.

Experimental assistants are briefed before the experiment starts about instructions, timing, checking answers, scoring, collecting messages, distributing questionnaires, and changing the screens for the second communication system.

Method. Working with six groups, three begin with the circle design, three with the wheel. Twenty-four subjects are assigned to their groups by random numbers and given a task to occupy their attention while others are being briefed as assistants. It is possible to work with more than one group in a room, if fairly well spread out. Five trials under each condition gives material for statistical comparison and a fair opportunity for learning to occur. This takes about twenty to forty minutes all told. Interest is well sustained, so that more trials can be given, but preferably not at the cost of discussion afterwards, which can give much additional information.

The subjects are given a general introduction to the effect that they will be participating in an experiment on communication, in which they will have problems to solve and will be able to communicate only by written messages. The rest of the instructions must be given at the

tables, where the flaps are open and closed as required for the first set of trials. For the circle design, communication with next-door neighbours is possible, but not across the table. The first instruction is as follows:

In this experiment on communication, you will not be able to talk to each other, but you have to solve problems by pooling information. This you do by sending messages to each other, written on the coloured cards. You may send messages through the windows marked IN only. But note, you may not use the windows when the flap is down – only those that are open.

In front of you there is a small pile of white cards turned face downwards. On each of these cards are drawn four symbols. The problem is to find out which symbol you all have in common – there will be one only. You do this by sending messages to those of the group with whom you can communicate by messages – messages of any kind. Every message you receive you must keep in your compartment. You must send out cards of your own colour only.

When you think you have the answer, be careful not to speak. Put up your hand and show me, when I come, the symbol drawn on a card. When *all four of you* are right, then I shall say so. I shall not say right or wrong to individuals.

At the end of the trial, the messages will be collected, so leave those you have received on the table in front of you till then.

Is that quite clear? Now turn over card no. 1 and begin. (The stopwatch is started.)

It is important to stress that all messages should be retained and not passed on: for analysis, every communication between the people must be on a separate card. The arrangement about not saying right and wrong to individuals is to match Leavitt's device of the master switch.

The time taken to reach a correct solution is recorded, together with a note of any incorrect ones offered, marked by the colour of the player. The messages are collected at the end of each trial. Each subsequent problem can be introduced by 'Now turn over to card X

and repeat'. After all the circle trials, questionnaires are distributed for everyone to fill in individually.

They are then told, 'I am going to change the flaps on the windows. Remember that you cannot use a window that is covered over.' To change to the wheel arrangement, the flaps between Pink and Yellow, and Pink and Blue, are closed, and the cross-bar flaps, between Pink and Green, opened. Thus everybody is in two-way communication with Green only. After the introduction, 'Here are some more problems. You do just the same as before, but using different windows. Now turn over to problem X', the wheel trials are run off, terminating with a second copy of the questionnaire.

Statistical treatment and results. A valuable aspect of this experiment is that a great deal of material comes out of it which can be distributed for analysis among the participants. At first, score sheets, messages, and questionnaires can be worked on separately, though after a certain stage cross-reference must be made to answer the main problems.

The chief results on efficiency come from comparing, by the Wilcoxon Signed Rank Test, the two sets of times: errors are likely to be too few for statistical comparison and concentrated on a few trials only. Learning curves for the two structures illustrate how they compare from a tactical point of view. All this material should be examined together with the summaries of individual evaluations of efficiency, from question 7, related to the two structures. In addition, the messages give information about the tactics adopted, and the extent to which there was concern with tactics. To some extent, the sheer number of messages reflects efficiency, though as Leavitt points out, theoretically at least the number of messages sent can be negatively correlated with time taken in a group situation where more than one message is being sent in a time unit. With all this material, transfer effects need careful scrutiny: thus, a succession of comparisons on scores should be made, wheel first vs. wheel second, wheel first vs. circle first, and so on. In fact, especially if group differences are of any interest, an analysis of variance can be calculated. Subjective impressions of transfer effects should have been gathered during the discussion following the experiment. In analysing the

messages, an interesting question is to ask whether the fate of a general tactical suggestion, or a short-cut in problem-solving, is related to the type of organization and the centrality of the suggester.

Our question about leadership is easily – and usually dramatically – answered from question 3. In the wheel design, Green is almost unanimously selected, with honours evenly spread (or no nomination at all) in the circle pattern, even when it has followed the wheel. And the distribution of the number of messages by colour reflects this result.

The assessments of liking, on question 5, can be compared by the Wilcoxon Test, although the distribution of the votes may be more interesting. Here, a comparison must be made between the two designs, and between the leader and the others on the wheel design. (In the wheel, unlike the Y-shape and the chain used by Leavitt, all non-central figures are equally peripheral.) Enjoyment can be related also to degree of comprehension of the pattern, from question 4, and to the extent to which, in questions 6 and 7, the need for better organization or better understanding of the organization was stressed. Both messages and questionnaires are worth analysing for evidence of frustration, sarcasm, and aggression (towards the experimenter, possibly), the incidence of which can be related to design and position. And finally, transfer effects need to be examined on these points.

Summary and conclusions. The superiority in productivity of the wheel is not predictable by simple calculation: in both designs, a problem is solvable by six messages (excluding those needed to inform the experimenter of the result). In all sized groups, ideally $2(n-1)$ steps are involved, for two-way transmission of shared information to every member. There is a slight difference in total sum of distances, the aggregate of all the steps linking member with member (16 for the circle, 18 for the wheel), but this difference is in the opposite direction from the differences in efficiency; however, the task does not require that all members should interact directly.

The difference can best be understood by reference to Bavelas' concept of 'centrality'. The circle design could be as efficient were there no ambiguity about who was the central figure, the collector of information. Thus the circle groups are much more confused about

their tactics and seldom settle down quickly to a steady plan. It is not surprising, then, that the central person in the wheel design should be acknowledged as the leader of the group: doubtless, were a circle group to settle down to an economical procedure, a leader would automatically arise there also.

The efficiency of the wheel design is apparently at the cost of the enjoyment of the members. Where there is one central person carrying most of the responsibility, he enjoys himself but the subordinates are relatively dissatisfied.

The importance of centrality to efficiency and liking can be shown most clearly, of course, when are included the two other designs, the Y and the chain, which contain intermediary centrality indices (calculated by Bavelas, from the relative sums of differences).

Practical applications. The recognition of one's role as relatively independent must be a contributory factor to its popularity. On the other hand, frustration with the inefficiency of the more democratic structure is also present. One might expect to find cultural differences here: the effect of independence was quite marked with Leavitt's American student subjects – a population, we are often told, very conscious of the importance of autonomy and control. Further research questions with practical relevance arise here. At what point, and in what circumstances, does frustration at inefficiency overtake enjoyment of equality and cause overall dissatisfaction with the circle pattern? Probably this will happen when the task is made more difficult, when the rewards and penalties are greater, and when there is greater personal involvement.

This experiment has fairly clear relevance to practical situations in which communication networks have to be established. The challenging question seems to be what pays best in the long run, high productivity or the satisfaction of the participants? The answer to this question may change an organization chart.

GROUP SIZE AND PROBLEM-SOLVING

Introduction. The question – When is it worth while to form a group? suggests the rider – and how big a group shall it be? Many

experiments have been conducted on this problem: these are well summarized by Argyle (1957) and by Kelley and Thibaut (1954). One cannot expect, of course, that experiments will produce a simple generalization about group size that will be universally applicable. The answer will vary according to the group's purpose and to the amount and kind of formal organization it is to have, as well as to more personal variables such as previous experience of working in such groups. The interest in this research area lies in tracing which antecedent variables, like these, are relevant and just how changes in size alter the dynamics of group interaction. Thus, an experiment in group size, relatively simple in terms of the experimental variables to be manipulated, can be very rewarding as an object of observational study.

The empirical problem. An early worker on group size, South (1927) contrasted groups of three and groups of six on problems for which a group decision had to be reached. Working with several types of problem, he concluded that larger groups have the advantage, when hypotheses have to be proliferated and tested, and smaller groups when working with material which lends itself to the immediate formation of opinion. The advantage of number in the kind of problem where bright ideas are needed is not only that there are more people to produce them but also more to criticize and reject. Specialization of function, even, becomes possible.

More recently, Taylor and Faust (1952, also in Hare, Borgatta and Bales (1955)) compared single individuals, groups of two and groups of four playing Twenty Questions. They found no statistically significant difference between the two group sizes, both of which were superior to individuals. Twenty Questions is a convenient game for such experiments. The rules are usually familiar by now, and it is not long before certain basic strategies are produced – such as asking – Human? quickly if it is Animal. Provided the groups are fully aware that they are free to plan their questions by discussion before trying them out, it becomes the sort of problem on which, on the basis of South's conclusion, one would expect a group of six to be superior. The limited design to be discussed here is a comparison of groups of three and groups of six. And as a compromise between

the demands of experimental rigour and the rewards of personal experience, all the eighteen subjects will have a taste of both groups.

Aim of the investigation. The main aim of the experiment, then, is to study the efficiency of groups of three and groups of six at solving a modified version of Twenty Questions problems: statistically, by comparison of time taken, number of questions asked, and of problems unsolved, where the hypothesis is that the larger groups will be more efficient; further, by discussion of the ways in which group size appeared to affect the work of the groups; and possibly, by subsequent analysis of tape recordings. The design suggested here also makes possible some examination of learning and transfer effects.

Apparatus. Four groups will be working at a time and each will need two or three experimenters furnished with a stopwatch, a list of problems (the same for all groups), and score sheets. A tape recorder for each group makes for added interest but this is not essential. At times, in fact, no use can be made of a recording if several people talk at once, and it is better to dispense with machinery than to inhibit the spontaneous activity of the group. The groups must not be within earshot as they are working on the same problems.

The problems, fifteen of which are needed for the present design, must be fairly straightforward since a high proportion of correct solutions is desirable. One way of simplifying the task is to use the categories Animal, Vegetable, Mineral only, unmixed. It is advisable to supply to experimenters notes for guidance on replies. They must also know the answers available to them: Yes. No. Partly. Sometimes. Not in the usual sense of the word. Please restate the question.

Also useful is a score sheet for each trial containing a seating plan against which can be marked the asker of each question.

Method. The design to be described here takes, in all, including briefing assistants and changing rooms between rounds, about one and threequarter hours. This is because a time limit of five minutes is set to each problem. That is necessary because, without it, endless procrastination may occur before the twentieth question is put.

Eighteen subjects are assigned to six groups of three by random numbers. The total experiment is divided into three rounds. In each,

two groups of three are working, the rest being paired as groups of six. Each person will work for one round in a group of three and for two in a group of six. Five problems constitute a round.

The subjects can be instructed before dispersing to their groups. They are told that they will be playing Twenty Questions in groups of different sizes, and what information they will receive about each article and the possible answers to their questions. For each problem a limit of twenty questions and five minutes will be imposed. Stress should be laid on the fact that they can talk freely to plan their questions: this must be stressed because it is not the usual practice in this game. They should also be reminded that they are not competing with each other within the group, but with other groups.

After the general introduction, nothing remains to the group experimenter but to present the first problem - This is - (animal, mineral or vegetable), starting the stopwatch simultaneously. Assistants look after the tape recorder, the record of questions asked and the record of questioners. Doubling-up is possible here, but the experimenter should be free to concentrate on answering questions as quickly and accurately as possible.

In the discussion that follows, it is useful to ask for predictions about the superior group size before the factual records are collated, and for suggestions about the factors that facilitate or impede successful working in the different groups. This discussion may suggest points to be tested in analysing the records.

Statistical treatment and results. Groups of people can work on the material requiring analysis. The main question about the relative efficiency of groups of six and three is answered by combining the results of all groups, making thirty results for each group size. The number of questions taken to solve the problem may be usable, given that there are not too many failures to complete. Taylor and Faust found it necessary to allow thirty questions and they gave a score of 31 questions for a problem unsolved. In comparing the groups by time taken, a similar arbitrary decision must be made for a failure, such as to score it as the full five minutes. These comparisons can be made by the Wilcoxon Signed Rank Test (or the t test for correlated scores if the distribution of results is suitable), and

the number of failures by the Chi-Square test. Little can be deduced about learning, since the problems have not been equated for difficulty. Transfer effects are concealed in the quantitative information, for the most part, but can be looked for in other material.

The records of the questions asked can be studied for evidence of learning. Criteria of good strategy will have to be worked out; for example, questions which are likely to reduce the possible field of choice considerably. Thus, the answer No to the question Is he in the entertainment world? is more informative than to the wild question (wild, that is, if it occurs at about question three) Is he in Twenty Questions on the wireless?

The records of questioners may show whether one particular spokesman emerges in a group and whether he remains the same in all rounds, especially whether it is the same person in groups of different size. (One analysis of this material showed a tendency for each threesome *within* a group of six to have one person who asked most questions, this not necessarily being the person who asked most when the threesome was by itself.) Transfer effects can be studied in the way in which questioning is distributed, since experience at being a threesome will occur in different sequential arrangements for the six groups.

The tape recordings, if available, can be very rewarding, but their analysis is time-consuming since listening to them once through will take, of course, four times as long as the original experiment. A group of workers each with a careful plan of action is needed. The search can be directed again at the contribution of individuals in groups of three and six, with consideration of transfer effects too, but now from the discussion that took place between the putting of questions. Are some individuals stimulated by greater numbers and others inhibited? And are those who seem to be the people with the best ideas susceptible to either of these pressures? Is it true that successful problem-solving is largely the work of one good person or is it the product of interaction on each other's ideas? And to what extent do increased numbers impede, by encouraging irrelevant and facetious discussion, or the repetition of questions? These are some of the questions that may guide the study of the tape recordings,

which should perhaps be done without knowledge of the quantitative findings.

Summary and conclusions. In bringing together the analyses of the various records, the main aim is to state which size of group has been more efficient, and what might seem to be explanations of the superiority. There are several possibilities to be looked at. Care must be taken that the answer is not quite simply that (in the case of the larger groups being better) one very good individual or trio performed for two rounds in a group of six and for only one in a threesome: this is a limitation of the experimental design, not to be remedied without more important sacrifice.

Another serious limitation to the design of the experiment is the fact that the experimenter who answers questions is not the same person in all groups. As one group of critics put it after being involved in the experiment, it is wrong to regard the team alone as the relevant social group: the team, the experimenter, and assistants are all interacting according to shared understandings.

Practical applications. In considering the applications of this experiment, and any pointers it might yield about the effects of increasing the size of problem-solving groups, care must be taken to consider the relative man-hours. A group of six may have greater output than a group of three – but there are twice as many of them to be paid for their time if they are a Works Council. In what conditions might it pay to have the larger group despite the extra man-hours? We are unlikely to find increase in size directly reflected in corresponding increase in output, as Gibb (1951) has shown. Perhaps, however, when the *quality* of the production is important, and where the problem area is large, more heads are better than fewer – to a limit, of course. Taylor and Faust say that the Twenty Questions situation has features resembling the activity of a scientific research team. A high level policy-making committee might be another case in point, or a jury obliged to produce a unanimous verdict.

Acknowledgements. Several points made here have been suggested by students, either in discussion after they have been experimental subjects, or in the reports they have composed on data analysed with

ingenuity and thoroughness. Thanks are due to them, and to the demonstrators, students, and friends who have assisted with the conduct of experiments at various times. Not least, the administrative and technical staff of my Department are to be thanked for their usual good offices.

REFERENCES

- ARGYLE, M. (1957). *The Scientific Study of Social Behaviour*. London: Methuen.
- BAVELAS, A. (1950). 'Communication patterns in task oriented groups'. *J. acoust. soc. Amer.*, **22**, 725-30.
- BION, W. R. (1948-51). 'Experiences in groups'. *Human Relations*, **1**, 314-20; **2**, 13-22, 295-303; **3**, 3-14, 395-402; **4**, 221-7.
- BROWN, R. W. (1954). 'Mass phenomena'. In Lindzey, G.: *Handbook of Social Psychology*, Vol. 2. Cambridge, Mass.: Addison-Wesley.
- CARTWRIGHT, D., and ZANDER, A. (1953). *Group Dynamics*. Illinois: Row, Peterson.
- FOULKES, S. H., and ANTHONY, E. J. (1957). *Group Psychotherapy*. Harmondsworth: Penguins.
- GIBB, J. R. (1951). 'The effects of group size and threat reduction upon creativity in a problem solving situation'. *Amer. J. Psychol.* **6**, 324.
- HARE, A. P., BORGATTA, E. F., and BALES, R. F. (1955). *Small Groups*. New York: Knopf.
- HEISE, G. A., and MILLER, G. A. (1951). 'Problem solving by small groups using various communication nets'. *J. abnorm. soc. Psychol.*, **46**, 327-36.
- HEYNS, R. W., and LIPPITT, R. (1954). 'Systematic observational techniques'. In Lindzey, G.: *Handbook of Social Psychology*.
- HOMANS, G. (1951). *The Human Group*. London: Routledge & Kegan Paul.
- ISAACS, S. (1933). *Social Development in Young Children*. London: Routledge & Kegan Paul.
- KELLEY, H. H., and THIBAUT, J. W. (1954). 'Experimental studies

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

- of group problem solving and process'. In Lindzey, G.: *Handbook of Social Psychology*.
- KLEIN, J. (1956). *The Study of Groups*. London: Routledge & Kegan Paul.
- LEAVITT, H. J. (1951). 'Some effects of certain communication patterns on group performance'. *J. abnorm. soc. Psychol.*, 46, 38-50.
- LINDZEY, G., and BORGATTA, E. F. (1954). 'Sociometric measurement'. In Lindzey, G.: *Handbook of Social Psychology*.
- MACCOBY, E. E., NEWCOMB, T. M., and HARTLEY, E. L. (1958). *Readings in Social Psychology*. New York: Holt.
- MINTZ, A. (1951). 'Non-adaptive group behaviour'. *J. abnorm. soc. Psychol.*, 46, 150-9.
- PIAGET, J. (1932). *Moral Judgment of the Child*. London: Kegan Paul.
- SOUTH, E. B. (1927). 'Some psychological aspects of committee work'. *J. appl. Psychol.*, 11, 346-68, 437-64.
- SPROTT, W. J. H. (1958). *Human Groups*. Harmondsworth: Penguins.
- TAYLOR, D. W., and FAUST, W. L. (1952). 'Twenty questions: efficiency in problem-solving as a function of group size'. *J. exp. Psychol.*, 44, 360-8.
- WHYTE, W. F. (1943). *Street Corner Society*. Chicago: Univ. of Chicago Press.

Chapter 5

INTERVIEWS

D. R. PRICE-WILLIAMS

London School of Economics



An interview can be distinguished from a conversation by defining it as a conversation with a purpose. Interviews are enacted in order that some information is gained. An interview is *for* something, not just an occasion for pleasant chit-chat. Somewhat more formally an interview has been defined by Kahn and Cannell (1957) as a 'specialized pattern of verbal interaction – initiated for a specific purpose, and focused on some specific content area, with consequent elimination of extraneous material'. To the careful observer and especially to anyone that has had experience of interviews, the last two passages of that definition are most important. The problem in an interview is not always so much the difficulty of getting somebody to talk, as keeping him on the subject once he starts. For most people are not unlike the character about which a nineteenth-century minor poet once wrote, whose

*... talk was like a stream, which runs
With rapid change from rocks to roses:
It slipped from politics to puns,
It passed from Mahomet to Moses ...*

An interview then has a purpose and has a definite subject matter.

Uses. Interviews can be classified according to the particular purpose which they serve. There are medical interviews, social work interviews, interviews for selection of candidates, political interviews, and so on. In each of these the intention of the interviewer is to gather information from those interviewed, whether it is of a

purely factual nature or whether, as might be the case with a medical interview, it is an attitude that is wanted. The use of the interview determines partly the form in which it is conducted. If all that is wanted is the attitude of the person being interviewed – the respondent – then rather formal questions with definite choice of answers are to be discouraged, as the purpose will be for the respondent to express himself at length and in his own way. On the other hand if the purpose is to gain information upon which a definite hypothesis stands or falls, then a fairly rigid approach is required. This leads to a classification of interviews which turns on the way in which questions are asked.

Types. The classification made by E. E. and N. Maccoby (1954) is based on a continuum which they call standardized-unstandardized, but which is similar to that called structured-unstructured by other authorities. The difference is well expressed by the Maccobys:

By a standardized interview we mean one in which the questions have been decided upon in advance of the interview, and are asked with the same wording and in the same order for all respondents. The questions used in a standardized interview may be either 'open' or 'closed'. The essential feature of the standardized interview is that the interviewer does not have freedom to word or reword questions, to introduce questions which seem especially applicable to the individual case, or to change the order of topics to conform to the respondent's spontaneous sequence of ideas. In the unstandardized interview, the interviewer's technique is completely flexible, and is varied from one respondent to another.

Varieties of interviews lie between these two poles. For example Merton and Kendall (1946) have introduced the 'focused interview' which has a fixed framework of questions yet allows the interviewer a certain amount of latitude within it. Again, interviews may start out completely unstandardized and become semi-standardized half-way through. A great many varieties are possible; the dimension is useful in being able to identify them.

Co-operation and 'rapport'. It may seem axiomatic that the motivation and co-operation of the respondent should be maintained

and that the relationship between the latter and the interviewer be agreeable, but these are points which are very often forgotten in the working out of the detailed technical aspect of the interview. One method of inviting co-operation is that of allowing the respondent to think that his views are important to society at large. Most people like to entertain the notion that their own ideas and habits are rules which others ought to follow, and gladly welcome the opportunity to air these to an interested listener. This is a natural tendency which can be exploited during the preliminary negotiations of the interview. Nothing unethical is being suggested, merely that the respondent should be allowed to feel that his responses are not being made in a vacuum, that while his *name* may be kept discreet, his *views* are placed in the forefront. For, as Kahn and Cannell have pointed out in Chapter 3 of their book already referred to, 'we give information to the doctor in order to get well, to the personnel interviewer in order to be considered for a job we want, to the social worker in order to get counsel or economic assistance, and so on'. Therefore, as they go on to say, 'When the interview begins, the interviewer must depend almost entirely upon making it meaningful to the respondent in terms of his needs and goals.'

'Rapport' is linked naturally with co-operation. Very little is likely to be gained if the interview is not conducted in terms of gaining the interest and sympathy of the respondent. Nevertheless it should be noted that 'rapport' can be pushed too far, as has been well emphasized by Hyman (1954):

We have assumed that great rapport and friendship patterns and a lot of social interaction are *requirements* for good interviewing, without ever observing the precise operation of those factors upon the behaviour of a respondent. Carried away by the emphasis on rapport, we have perhaps vulgarized the concept and have mistaken 'love' for rapport. And interviewers may have followed suit, and striven for great chumminess with their respondents. A certain degree of businesslike formality, of social detachment, may be preferable. When rapport transcends a certain point, the relationship may be too intimate, and the respondent may be eager

to defer to the interviewer's sentiments. This would seem especially the case when the respondent has little real involvement in the task. When he is not particularly interested in the issues or has no strong views of his own, he may not mind or even prefer to take over the coloration of a very friendly interviewer. Perhaps, when the issues are of such a character as to create real task involvement, there is a counterbalance to the deleterious effects of excessive rapport.

Sources of error. What is meant by error in interviewing turns chiefly on the use of the interview. In a clinical interview, when all that is required is for the respondent to talk about himself and his attitudes, error lies in the general reliability in what he is saying. If interviews are intended for selection purposes, which they very often are, then error lies in the validity of the proceedings, that is whether the interview has been successful in selecting the type of candidate that was wanted. In straightforward gathering of information on factual issues, which is the type of interview that we are considering here, error lies very largely with the expectations of the respondent in relation to the interviewer himself and the type of question that is asked, points which will now be discussed.

(1) *Interviewer effect.* A general tendency among respondents is to answer more freely to interviewers whom they look upon as being like themselves and to be restrained to interviewers who are looked upon as being unrepresentative of their own particular group (Newcomb, 1950). An early experiment to indicate this tendency was done by Katz (1942) who showed that you could get differences in the responses of people who were interviewed by working-class interviewers and by white-collar interviewers. Cantril (1944) followed this study up by reporting differences in attitudes of Negro respondents towards a particular policy of the Axis powers according as to whether Negroes or Whites were the interviewers. In both these cases the source of error was the appearance of the interviewer. Besides this the effects of the interviewer's own attitudes and expectations have a role. A large amount of studies now exist to show pretty convincingly that the role of the interviewer is vital in any assessment

as to the reliability and validity of the interview. Hyman and his associates (1954) provide the larger part of the evidence on this score and go on to suggest measures for correction of the factors involved.

(2) *Invalidity of the interview.* There are a number of studies which show that the factual information offered by the respondent is incorrect or lacking in some way. A study by Parry and Crossley (1950) who asked a large number of people in one community questions about their age, ownership of library card, and of driver's licence, whether they owned a telephone and other similar enquiries, and who then checked up on these responses from other sources, showed that a high percentage of these responses was inaccurate. Again when people are asked certain questions and then re-interviewed some time later with the same set of questions, there are found differences. Kinsey, in gathering data for his book on sexual behaviour in the male (1948), reported such discrepancies. This is an issue which the first of the experiments to be described will take up directly.

(3) *Question content.* Particularly in an unstandardized interview, the way in which a question is put colours the form of response. This, after all, has been known for some time in law courts. We come back to the attitudes of the interviewer, because his own viewpoint very often determines the actual question content. The point is perhaps too obvious to labour with examples from studies, but the process is sometimes so insidious that it is worth while quoting Kornhauser's (1951) guide list for formulating questions. It must be borne in mind that it is not only the attitude of the interviewer that may influence the question content – the actual formulation of questions has its own mechanics which are not free from error. Kornhauser's list is as follows:

- Is this question necessary? Just how will it be useful?
- Are several questions needed on the subject matter of this question?
- Do respondents have the information necessary to answer the question?
- Does the question need to be more concrete, specific, and closely related to the respondent's personal experience?

Is the question content sufficiently general and free from spurious concreteness and specificity?

Is the question content biased or loaded in one direction – without accompanying questions to balance the emphasis?

Will the respondents give the information that is asked for?

Each of these items contains sub-items which ask further questions as a guide, and the student of interviews should consult them before hurling himself into the fray of interaction with a respondent, but the above quoted list should indicate the common pitfalls. The factor of question *content* is quite apart from that of *wording*, which will be the basis of the third experiment, and about which more will be said later.

(4) *Recording*. There are at least four methods of taking down what the respondent says, all of which can lead to some kind of error. The first is merely writing-up the interview from memory. This suffers from so obvious demerits that it is unnecessary to pursue it further. All that we know from the psychology of perception and remembering generally, will lead us to suppose that the recorded interview will very largely be a reconstruction (Bartlett, 1932).

The second method, the one provided in the set of experiments here, is that of pre-coding. The respondent is provided with a list of alternative answers, which can be given a number, and all that the interviewer has to do is to check it. This is more suitable when closed questions are asked rather than open ones, although even in the latter case pre-coding is possible. The limitation of this method and the possibility for error lie in the fact that when open questions are pre-coded there is a temptation by the interviewer to pigeon-hole the responses into his pre-coded answers, a discovery made by Woodward and Delott (1952). The third method is taking down verbatim what the respondent is saying, while the fourth method merely substitutes the tape recorder for the interviewer's shorthand. If everything is faithfully recorded and the interviewer has the time to go over the notes made, whether mechanically or otherwise, there is little chance for error to creep in. Practical experience, though, has shown that verbatim note-taking very often does *not* take down everything, is very laborious and time-consuming in transcription, while tape

recordings can be indistinct, are often off-putting to the respondent, and again take up time in playing through afterwards. The merits and demerits of each of these forms of recordings are relative to what is demanded. For a completely unstandardized interview with free open questions, it may be advisable to use a tape recorder in view of the difficulty of making any pre-coding. On the other hand when all that is required is information and the form of answer is closed, then it would be unnecessary and wasteful to use any mechanical means.

The empirical problem. It is with the sources of error that the following three experiments are really concerned. First of all each experiment has two interviews, so that the re-interview effect (see Rosenberg *et al.*, 1951) is studied throughout, although this problem is focused upon in the first experiment only. This problem has tended to arise, as Rosenberg points out, in Panel Studies where repeated interviewing 'leads respondents to change their answers in some way so that they no longer represent the opinions and attitudes of the population they are intended to sample'. In these experiments more than two sessions is impracticable, so it is suggested that this should be the maximum with a minimum time limit between them of one week. One month, if possible, is a better duration.

The question of the sex of interviewer in relation to the sex of the respondent is a matter which has received some attention (see Chapter IV of Hyman and others, 1954). It seems plausible that, for example, opinions about sexual behaviour might be a subject concerning which the sex of the interviewer may have an effect. Experiments along this line so far do not appear to indicate that common-sense assumptions are vindicated but more subtle differences may well be disclosed. At any rate this is a factor in interviewing which is open to far more research.

The wording factor is so basic that practically entire books have been devoted to it. The reception given to various words seem to have so many nuances that it is difficult to find a term which does not arouse associations and connotations which are unintended, and thus it is a perennial problem in the art and science of interviewing.

DESIGN OF THE EXPERIMENTS

The research design which is now going to be described in detail has the advantage that the same interview schedule can be used in three different conditions, which therefore allows for three different experiments. The three conditions are (1) The Re-interview factor, (2) The Sex factor, and (3) The Wording factor. In the first case the problem is raised as to the reliability of an interview. Will a person who has answered a list of questions in one way identically repeat those answers if given the same questions after a period of time? And will this same person, on re-interview, repeat his previous answers if he is now interviewed by a different interviewer? These are questions which the first experiment will test. The second experiment raises the question whether being interviewed by a member of the opposite sex makes a difference in the responses. And if so, what kind of question is most influenced by this factor? Although these two experiments can be carried out separately, the design is such that there is a certain overlapping which makes for inter-comparison, which is later described in a separate section. Whereas experiments I and II employ the same interview schedule, which we will call Schedule A, the third experiment introduces variation of wording to the questions, which we will call Schedule B. The prime question in this case is whether a slight change in wording alters the nature of the responses, either in an expected or in an unexpected direction. As both male and female interviewers are also employed in Experiment III, again there is possible an inter-comparison with the first two experiments.

Choice of respondents. The design illustrated in Table I (p. 143) allows for seven male interviewers and seven female interviewers. These fourteen interviewers, taken together, interview fifty respondents. Specifically, each interviewer, or when the case arises, pair of interviewers, is responsible for five respondents only. The problem then arises as to the choice of these respondents. It is probably easier to choose respondents of the same sex throughout the three experiments, although it would be quite possible to arrange for the two sexes to be interviewed provided that the basic principle of the

design is maintained. Nevertheless in this design we will assume that the respondents are either all men or all women. The choice of sex will be determined by the ready availability of either sex.

Keeping the sex constant, ideally we should further select respondents within each experiment such that each set of five are comparable in age, intelligence, social class, and perhaps other attributes which we consider might make for differences in an interview situation. These make for rather stringent criteria which may not always be possible to follow. Accordingly we can follow the ruling that there should not be too much variation in subjects. For example 'young men' or 'young women' could be selected for each or all of the three experiments, or alternatively 'old men' or 'old women'. It will probably be advisable to avoid large variation in social class or occupation. For example the occupational choice of respondents for each experiment could be all students, or all unskilled workers. The actual suggested list of questions given here is biased in the direction of students of an age range from eighteen to, say, twenty-three, as can be inferred from the type of question, but there is no reason why questions could not be adapted to different age levels and to different occupational classes.

There is one further point to be considered in the choice of respondents, and that is the degree of familiarity with the interviewer. It is a far better test to choose people that one is totally unfamiliar with than to choose friends or acquaintances. For example students in a University should seek for their subjects those students taking another subject entirely, so that the degree of familiarity is minimal.*

The interview schedule. The type of enquiry around which the questions are aimed might be labelled informally as 'An Investigation into Some Common Vices', or more decorously 'An Investigation into Some Social Habits'. The investigation concerns such habits as

* These are the criteria for choice of respondents. In situations where it is possible and preferable, an instructor or lecturer can take over the whole business of choosing subjects himself, and also arrange the interviews to be done in class time. Re-interview arrangements would also be done by him.

smoking and drinking, and somewhat more drastically, misdemeanours such as avoiding payment on public transport. The subject matter has been chosen on account of its inherently motivational nature on the one hand, and on the other hand for its sensitivity towards differing conditions in the interview situation, which are the point of the experiments. In other words, with these experiments we are not so much interested in *what* people say as the way in which their answers are altered by varying the interviews and the interviewers.

Below is a list of questions, which represent Schedule A. The altered text, which we will call Schedule B, will be given when the experiment that concerns i. is presented in detail.

QUESTIONS

1. Do you smoke?
2. Do you drink alcohol?
3. Do you attend wild parties?
4. Do you get out of bed late?
5. Are you late for appointments?
6. Do you keep your belongings neatly?
7. Do you wash yourself thoroughly?
8. Do you take girls out?
9. Do you avoid payment on public transport?
10. Do you lie to people?
11. Do you get into debt?
12. Do you talk badly of others?

It can be seen that the 'vices'* concerned are, among others, to do with laziness, unpunctuality, dishonesty, maliciousness, uncleanness, and lack of temperance and of personal system. They are not so extreme that people would be embarrassed to answer nor so slight

* Care must be taken not to assume that the 'vices' in question generalize in a given personality, e.g. that a man apt to be late for breakfast is also liable to be late for business appointments. It should not be assumed that there is a 'vice' for unpunctuality in general.

that a variation in answers would not occur. If new questions are to be substituted in order to fit in with a different population, these criteria should be kept in mind. Thus if elderly people are to be chosen as subjects, Question No. 3 would not be a good one. It is unlikely, though one never knows, that old men or old women would attend wild parties, at least with a regularity that would make such a question a relevant one. Also to be noted is the fact that the questions are arrayed in an order of mild to bad 'vices', assuming a usual set of values. This has the advantage that one leads in the respondent from a relatively innocuous set of questions to those which might cause slight embarrassment.

Coding of Responses. The questions are 'closed' to the extent of there being five possible responses, as follows:

0. Never
1. Seldom
2. Sometimes
3. Often
4. Very Often

Measurement can be done on this five-point scale, but in addition to the fixed responses, it is useful to have a qualitative 'open-end' answer to each question in case a respondent wishes to qualify a response. For example he may want to qualify a 'sometimes' response to the question 'Are you late for appointments?' by saying 'not intentionally'. Or he may admit to lying 'sometimes' with the addendum of 'when necessary'.

*Pilot trials with the schedule.** Questions must be clear and free from ambiguity. In order to find out whether a chosen set of questions is clear, a small group of subjects, similar to those later to be selected as respondents, can be tried out with the proposed schedule. Here the experimenter is interested not in the answers but only in whether the questions are understood. This is really an experiment

* Here again arrangements can be determined in advance by an instructor. Although it is instructive for students to go through this particular stage, as far as they are concerned the pilot trials can be assumed to have taken place.

in itself, and will certainly throw up points which the experimenter is unlikely to have anticipated. For example many people will require some clarification of what is meant by a 'wild' party. Some will want to know what is 'late' in Question No. 4. Other difficulties are less obvious. A respondent may not be sure whether owing money to a bank counts as getting into debt. Others may hedge at the 'washing' question on the word 'thoroughly': does this mean washing the hair also? Again, the question about taking girls out will run into difficulties when the respondent is married. Such clarification points can be numerous, and require that pilot trials are first done, and that the interview schedule proper is only presented after amendments following from these trials. Therefore the list of questions presented here is put forward as a model only, prior to piloting, so that the difficulties inherent in asking questions can be experienced and worked through. It is a common feature of interviewing that, apparently, the simplest question can be misunderstood. It is most instructive that this should be experienced.

General design of the experiments. This is given in Table I. The convention has been followed that the sequence of male interviewers are identified by letters of the alphabet beginning at the end, that is with Z, and that female interviewers begin with the letter A. All the respondents are in ordinary numerical sequence.

This design can be written out in full in a notebook. As the different combinations can be complicated, it is essential to have the full plan in front of one, in order that the results can be assessed properly.

Certain points can be seen from this Table. Each interviewer, male or female, has a total number of five interviews to do. The exceptions to this are interviewers Z and A, who have ten, in Experiment I, and interviewers U, T, F, G, in Experiment III, who also have ten. The following interviewers work in pairs, seeing the same respondents: Y and X, B and C, W and D, E and V. Lastly, each interviewer, or pair of interviewers, will be responsible for finding five respondents.

Further matters concerning the design will be explained under the heading of each experiment.

General instructions. In order to avoid social difficulties, approach

INTERVIEWS

TABLE I. Design Framework

Experiment I. Re-Interview Factor

- | | | |
|--------------------------|---------------|--------------------|
| (1) Male Interviewer Z | | |
| interviews respondents | I, 2, 3, 4, 5 | |
| re-interviews | „ | I, 2, 3, 4, 5 |
| (2) Female Interviewer A | | |
| interviews | „ | 6, 7, 8, 9, 10 |
| re-interviews | „ | 6, 7, 8, 9, 10 |
| (3) Male Interviewer Y | | |
| interviews | „ | 11, 12, 13, 14, 15 |
| Male interviewer X | „ | |
| re-interviews | „ | 11, 12, 13, 14, 15 |
| (4) Female Interviewer B | | |
| interviews | „ | 16, 17, 18, 19, 20 |
| Female Interviewer C | | |
| re-interviews | „ | 16, 17, 18, 19, 20 |

Experiment II. Sex Factor

- | | | |
|------------------------------|--------------------|--------------------|
| (1) Male Interviewer W | | |
| first interviews respondents | 21, 22, 23, 24, 25 | |
| Female Interviewer D | | |
| re-interviews | „ | 21, 22, 23, 24, 25 |
| (2) Female Interviewer E | | |
| first interviews | „ | 26, 27, 28, 29, 30 |
| Male Interviewer V | | |
| re-interviews | „ | 26, 27, 28, 29, 30 |

Experiment III. Wording Factor

- | | | |
|--|--------------------|--------------------|
| (1) Male Interviewer U | | |
| interviews with Schedule A respondents | 31, 32, 33, 34, 35 | |
| re-interviews with Schedule B | „ | 31, 32, 33, 34, 35 |
| (2) Male Interviewer T | | |
| interviews with Schedule B | „ | 36, 37, 38, 39, 40 |
| re-interviews with Schedule A | „ | 36, 37, 38, 39, 40 |
| (3) Female Interviewer F | | |
| interviews with Schedule A | „ | 41, 42, 43, 44, 45 |
| re-interviews with Schedule B | „ | 41, 42, 43, 44, 45 |
| (4) Female Interviewer G | | |
| interviews with Schedule B | „ | 46, 47, 48, 49, 50 |
| re-interviews with Schedule A | „ | 46, 47, 48, 49, 50 |

to respondents is best done by referring to the fact that this is not an official enquiry of any kind but merely an exercise for students of interview-techniques. This at once disarms apprehension of disclosing private foibles and facilitates the required interviewer-respondent relations. People are put into a situation where they feel they are helping out a student. Nevertheless the type of question asked has just that little bit of delicacy where an approach to a respondent should be tactful. The latter can be told that he need not answer if he does not wish to. Again, anonymity can be preserved; the respondent's name need not be taken. A code number is all that is required. Further specific instructions will again be given with each experiment. This can be made more formal by providing each subject with typed instructions pasted on cardboard. This has the advantage of standardizing the instructions throughout the experiments. After the first approach, the experimenter may say: 'With your permission I am asking you to read the instructions for this experiment; if it is not clear, please ask questions.' The standard typescript will then be given to the subject and runs as follows:

This is not an official enquiry. NO FURTHER USE WILL BE MADE of any personal fact contained in your answers, which will be treated with full professional secrecy. No personal details will be allowed to leak out, nor will any disclosure of them be made TO ANYBODY. This is an exercise for students of interviewing techniques only. Please answer as truthfully as you can, though you may find that some of the questions come rather close to the bone. NOT EVEN YOUR NAME WILL BE RECORDED, ONLY A CODED NUMBER.

After he has read this, the subject can be told that if he wishes to, the results of the enquiry will be told him when it is all completed.

THE RE-INTERVIEW FACTOR

Aim of the investigation. This experiment deals with the problem of reliability of response. No doubt the reader will have seen reports of interviews, both by journalists and by psychologists, and wondered

whether answers to the questions were haphazard or whether they really reflected genuine information. Is one likely to say 'No' to one question on Monday and to say 'Yes' on Saturday? Does, in fact, the mere occurrence of being re-interviewed over identical material cause a respondent to alter his previous answers? These are questions which can be put to the test in the present experiment. It is possible to go further than this and put forward certain assumptions. We might predict that a person will repeat his answers to those questions in the schedule which can be labelled relatively innocuous 'vices', but will tend to shift when rather more serious habits are investigated. Again, a prediction might be made that a respondent will repeat what he has said to the *same* interviewer but alter with a *different* interviewer. The design, as can be seen from Table I, allows for all these questions and predictions to be tested. There are other predictions which could be made, but let us take the above two and put them in more formal terms. As we are dealing with a relative scale of response we shall have to go further and predict the direction of shift.

Prediction One. Either with the same or with a different interviewer, answers to Questions 1 to 8 ('innocuous habits') will tend towards duplication of previous answers; those to Questions 9 to 12 towards modification in the direction of 'less culpability'.

Prediction Two. In the re-interview situation there will be a wider range of response with a different interviewer than with the same interviewer. Notice that the prediction does not say which way a shift may occur, merely that the range of response is less narrow.

Prediction One is based on the underlying assumption that a second interview on questions which probe somewhat serious 'bad habits' will tend to minimize scores, no matter with the same or a different interviewer. Prediction Two takes all the questions together and assumes that variability of response is more likely to take place with another interviewer than with the same one. These two predictions are to be taken independently: the first is concerned with *content*, the second with *consistency*.

Instructions and method. The general instructions are followed as outlined previously, in dealing with a respondent. After 'rapport'

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

has been established and preliminary talk completed, the experimenter asks the subject if he will answer twelve questions in terms of the fixed answers provided. At this point the subject is given a little card with the scale of responses, without the figure coding. The respondent may also be told that if he wishes to qualify these responses or make any comment, he may do so. The investigator will have a sheet, shown as Table 2, on which the answers are marked.

TABLE 2. Recording Responses
(See Coding of responses, p. 141)

<i>Interviewer: Z</i>		<i>Experiment I</i>		<i>Date: 3/6/61</i>		
<i>Respondent: 3</i>		<i>First Interview.</i>		<i>Time: 2.30-2.45</i>		
<i>Q.</i>	<i>Response</i>				<i>Remarks</i>	
(1)	o	I	<input checked="" type="checkbox"/> 2	3	4	
(2)	o	I	2	<input checked="" type="checkbox"/> 3	4	
(3)	o	<input checked="" type="checkbox"/> I	2	3	4	Only in holiday-time
(4)	o	I	<input checked="" type="checkbox"/> 2	3	4	
(5)	o	I	<input checked="" type="checkbox"/> 2	3	4	
(6)	o	<input checked="" type="checkbox"/> I	2	3	4	
(7)	o	I	2	3	<input checked="" type="checkbox"/> 4	
(8)	o	I	<input checked="" type="checkbox"/> 2	3	4	
(9)	<input checked="" type="checkbox"/> o	I	2	3	4	
(10)	<input checked="" type="checkbox"/> o	I	2	3	4	Never consciously, that is
(11)	o	<input checked="" type="checkbox"/> I	2	3	4	
(12)	o	<input checked="" type="checkbox"/> I	2	3	4	Not without cause

A square around the response number code indicates a respondent's answer.

As, in this case, two interviews are intended, it should be made clear to the subject in advance that two sessions are intended, each lasting not more than twenty minutes. At the beginning of the

second interview the subject is shown the following instructions: 'I should be grateful if you would answer the questions I am going to read. You will recognize them as those you answered before, but I want to get some further information. Please remember that your answers will be completely secret and anonymous.' In the case of re-interview with a different interviewer, the latter merely says that he is there in place of his friend, the original interviewer, and make no comment. If he is pressed by the subject he can say that his friend was unable to come. Thenceforth the new interviewer proceeds as described.

Analysis of results. In order that one can glance quickly at the overall results, averages can be computed for each question for each set of five subjects. In that part of the experiment in which the same interviewer re-interviews the same five respondents, we might get the following result:

<i>Q.</i>	<i>1st Interview</i>	<i>Re-interview</i>
(1)	2.4	2.6
(2)	1.6	1.8
(3)	4.0	3.8

These results are in averages, but the scores can be converted in different ways so as to allow for categories of response. Prediction One splits up into two halves. The first half says that answers on Questions 1 to 8 tend towards sameness of response either with the same or with a different interviewer. By taking these questions as one set and the remaining four questions as a second set, and comparing these as to whether the responses are the same or different, we obtain a 2×2 contingency table, on which the Fisher Exact Probability test can be calculated. The second half of the prediction says that on Questions 9 to 12, there will be a tendency towards reduction of response. Again the procedure can be used if 'Downwards' is taken as one category, and the same response plus any upward change are lumped together as the second category.

The chi-square can be further used for the second prediction. Here we are only interested in the change of *range* of responses. In each group the differences of responses between the first interview

and the second can be subtracted. As we are not interested in the direction, pluses or minuses are disregarded. This leaves a range of differences from zero to a possible 4. It was predicted that in one group the range of differences would be wider than the other. The cut-off point, designating 'wider', is the median, which in the range that we have here is 2.0. Thus a 2×2 table can be set up, the two classes of responses being scores above and below that median. The statistical test follows as usual.

Conclusions. The predictions stand or fall according to the findings of the statistical techniques. If they should fall it is a useful plan to re-consider the primary assumptions and look at the data from another point of view. The qualitative comments made by respondents should be inspected, especially to see whether they provide any indication for a further hypothesis.

THE SEX FACTOR

Aim. This experiment involves a second interview, but we are only interested in the probable effect of the sex of the interviewer. That is why, in the design, the serial order of interviewers has been taken into account. Consider the design again, lettering each section:

(a) Male interviewer	W	interviews respondents,	21, 22, 23, 24, 25
(b) Female	„	D re-interviews	„ 21, 22, 23, 24, 25
(c) Female	„	E interviews	„ 26, 27, 28, 29, 30
(d) Male	„	V re-interviews	„ 26, 27, 28, 29, 30

From this we can directly compare (a) and (c), because we have taken care of the order of sex by including (b) and (d). On a common-sense assumption (which can always be completely wrong) a hypothesis could be made that men, when being interviewed by women, will tend to tone down their responses on such a subject matter as this. On the other hand, on an equally common-sense assumption, one might postulate the opposite tendency, at least with some questions. It depends upon the attitudes which we think the respondents may have. If the attitude is that of a 'front' of respectability, then the first hypothesis will hold. If the attitude is that of boasting, then the second assumption would be correct. When one is faced with

the difficulty of not knowing which attitude is operating, the safest plan is to merely compare the two sets of interviews and leave it at that. For this purpose we can compare (a) and (c) directly, and we can further compare (a) with (b), and (c) with (d), for the reliability factor connected with change of sex of interviewer.

Instructions and method. The instructions given to the respondents and the method of execution are exactly the same as in the previous experiment. When the second interviewer, male or female, appears, second-interview instructions will be used as before. Scoring is done in exactly the same way also, except of course that the headings will be different.

Analysis of results. As we are concerned with finding out whether interviewer-difference in sex alters responses, we proceed by investigating question by question. This has the merit of showing up any particular question if in fact there are differences. From this point further analysis could be made. A nonparametric test, the sign test, can be applied to the comparison of (a) with (c), of (b) with (d), and if required, (a) with (b), and (c) with (d). We can use this test if we treat the various respondents as pairs. Thus 21 and 26 can be counted as one pair, 22 and 27 as a second, and so forth. When re-interviews are done then the same respondent, taken twice, can be considered as a pair. In each of these cases the number of pairs is five. The values taken for noting the difference between the two treatments are the raw scores of the question. Table 3 gives the scoring design necessary.

TABLE 3. Scoring for Experiment II

<i>Respondents</i>	QUESTION FOUR	
	<i>Male Interviewer</i>	<i>Female interviewer + or -</i>
21 & 26		
22 & 27		
23 & 28		
24 & 29		
25 & 30		

In using this scoring system it must be remembered that higher scores represent a higher frequency of incidence. The table serves

for the first two comparisons we wish to make. For comparing (a) with (b), and (c) with (d), the respondent 'pair' is the same person. If only five pairs are used, we can only get values at the 10 per cent. level of significance, but if the serial order factor is chosen to be ignored, both the male parts of the experiment can be collected, making ten pairs. Similarly with the females.

Conclusions. As has been said previously, these results will indicate whether there are some questions which seem to be affected by the sex of the interviewer. Assessment of the experiment proceeds first by singling out those questions out of the twelve which seem so influenced, and then analysing them in order to see what element they have in common, if any. This will serve as a hypothesis for further experiments which any investigator can make up. The fundamental design for further experiments consists of two sets of questions, those which contain this element and those which do not.

THE WORDING FACTOR

It has long been known from psychological research and otherwise that the way in which a question is worded will profoundly influence the answer to it. Muscio (1916) for example in some experiments reported during the first World War showed how apparently relatively simple changes in question wording, from putting the question in the form of the negative instead of the affirmative, or from using the definite article instead of the indefinite article, will give different responses. Payne in his book *The Art of Asking Questions* (1951) gives a whole chapter of problem words which make trouble in one way or another in questions. One of them, the simple word 'ever', we shall employ in this experiment.

The interview schedule. There are three changes in wording to the original Schedule A. Questions 1 to 4 are altered by including the words 'how much' or 'how often' at the beginning of the question. This forces the respondent to focus attention on the quantitative aspect of the question directly. The second set of four questions (5 to 8) interpolate the word 'ever'. As Payne has said about this word, 'Ever is such a long time and so inclusive that it makes it

INTERVIEWS

seem plausible that some unimpressive things may have happened.' The last set of four, Questions 9 to 12, are introduced by a reference to 'some people'. Some People avoid payment on public transport. Do you? The intention here is fairly obvious. By inferring that other people, or *some* other people, do these things, there is some degree of mitigation in case *you* would also do them. At any rate this can be tried out as a hypothesis; few assumptions in this field can be taken for granted.

The altered list, which we now call Schedule B, is thus as follows:

QUESTIONS

1. How much do you smoke?
2. How much do you drink alcohol?
3. How often do you attend wild parties?
4. How often do you get out of bed late?
5. Are you ever late for appointments?
6. Do you ever keep your belongings neatly?
7. Do you ever wash yourself thoroughly?
8. Do you ever take girls out?
9. Some people avoid payment on public transport. Do you?
10. Some people lie to others. Do you?
11. Some people get into debt. Do you?
12. Some people talk badly of others. Do you?

Method and instructions. There is little difference in method from the first two experiments. There is a straightforward interview followed by a re-interview with the same interviewer. The design allows for the factor of serial order of the two schedules. In two cases B follows A, and in two cases A follows B.

Predictions. The design is so constructed that the sex factor and re-interview factor can be picked out, but this time the basic interest is the change of wording. For each change of wording in the three sets of four questions which contain it, three separate predictions can be made.

(1) Inclusion of the quantitative terms 'how much' or 'how often' will make for a narrowing of range of responses.

(2) The word 'ever' will serve to cause general admittance of more frequency.

(3) Reference to other people will also induce responses on the higher frequency side.

Analysis of results. The raw scores again are the averages. A suitable table for recording of responses is given below.

TABLE 4. Scoring for Experiment III

Q	A	B	$A(1)$	$A(2)$	$B(1)$	$B(2)$	$d A-B$
1-4							
5-8							
9-12							

The first two vertical columns will give the over-all results of both Schedules. The next four columns will show any difference due to the order in which the respondent has received the different schedules. The last column, in terms of pluses or minuses, will register whether the above predictions are correct or not, or at least the second and the third predictions. These should show, if the predictions are correct, a minus sign. The first prediction is in terms of *range* and *dispersion*, and the relevant formulae for these are the best indicators. Student's 't' test is suitable for the testing of whether the differing averages are significant or not.

Assessment. There is considerable practical knowledge to be gained from this experiment. If the predictions turn out to be correct, or if we find in some other way that inclusion of relatively simple words or phrases changes the degree of responses, then questions should be so pruned and analysed in order that no undue tainting is produced by the form of it. It is instructive to glance at the literature of the subject in relation to one's own experiments, and wonder perhaps whether it is ever possible to coin a query that is 'pure'. In assessing results, one point needs to be kept in mind, and that is the variety of *perceptions* of words and phrases in different people. It is useful to have a back-check of the effect of the altered wording from the free introspections of the subjects. For instance one lead is to follow up what the words 'Some people' mean to various respondents. If results are discrepant in some way, the source of the discrepancy may

lie in the interpretation of words. The issue is really basic to all interview construction and practice. We faced it at the pilot level and we face it now again.

RELATIONSHIP BETWEEN EXPERIMENTS

It is clear that there are two recurrent variables running right throughout the three experiments. These are sex and the re-interview situation. Across them are the further factors of the same versus a different interviewer, and, in the third experiment, an altered text. A re-orientation of the basic design of Table I will give information which runs across all the experiments. Thus it is possible to collate the collected sex factor, comparing the results of ten male interviewers with ten female interviewers, irrespective of whether there is a different interviewer on re-interview, and irrespective of change of questions. Likewise all the first interviews can be compared with all the second interviews, ignoring the sex of the interviewer and ignoring again the altered wording.

Acknowledgements. I should like to thank my colleague, Dr A. N. Oppenheim, for advice, and the Social Science Students at the London School of Economics, who tried out the set of experiments.

REFERENCES

- BARTLETT, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge Univ. Press.
- CANTRIL, H. (1944). *Gauging Public Opinion*. Princeton Univ. Press
- HYMAN, H. H., with COBB, W. J., FELDMAN, J. J., HART, C. W.,
STEMBER, C. H. (1954). *Interviewing in Social Research*. Chicago Press.
- KAHN, R. L., and CANNELL, C. F. (1957). *The Dynamics of Interviewing*. New York: John Wiley & Sons; London: Chapman & Hall.
- KATZ, D. (1942). 'Do interviewers bias poll results?' *Public Opinion Quart.*, 6, 248.
- KINSEY, A. C. (1948). *Sexual Behavior in the Human Male*. Philadelphia: W. B. Saunders.

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

- KORNHAUSER, A. (1951). 'Constructing questionnaires and interview schedules'. In Jahoda, M., Deutsch, M., Cook, S. W. (1951): *Research Methods in Social Relations*, 1st Edition, Part 2. New York: Dryden Press.
- MACCOBY, E. [E. and N. (1954). 'The interview: a tool of social science'. In Lindzey, G. (1954): *Handbook of Social Psychology*, Vol. 1. Cambridge, Mass.: Addison-Wesley Publishing Co. Inc.
- MERTON, R. K., and KENDALL, P. L. (1946). 'The focussed interview'. *Amer. J. Sociol.*, 51, 541; also with M. Fiske, *The Focussed Interview*, Bureau of Applied Social Research, Columbia University, New York (1952).
- MUSCIO, B. 'The influence of the form of a question', *Brit. J. Psychol.*, 8, 351 (1916).
- NEWCOMB, T. M. (1950). *Social Psychology*. New York: Dryden Press.
- PARRY, H., and CROSSLEY, H. 'Validity of responses to survey questions'. *Public Opinion Quart.*, 14, 61 (1950).
- PAYNE, S. L. (1951). *The Art of Asking Questions*. Princeton Univ. Press.
- ROSENBERG, M., THIELENS, W., with LAZARSFELD, P. L. (1951). 'The panel study'. In Jahoda, M., *et al.*, op. cit. (See Kornhauser, A., above.)
- WOODWARD, J. L., and DELOTT, J. (1952). 'Field coding versus office coding'. *Public Opinion Quart.*, 16, 432.

Chapter 6

SOCIAL CLASS

GUSTAV JAHODA

University of Glasgow

This chapter, unlike most of the others, is focused on a specific topic, and each of its sections outlines methods and procedures for studying a particular aspect of social class. Apart from the first, which has been specially devised and would be difficult to modify in practice, these methods are not necessarily tied to any particular type of content. It is thus possible to use this chapter merely as a source for methods that can be applied in other fields.

The arrangement of the sections could be said to follow a sequence from quick and narrow to slow and wide. A short version of the first experiment can be carried out with student subjects according to detailed instructions and takes little time. The second experiment requires more careful preparation and makes greater demands on those conducting it, although the procedure is also fully described. The third section just offers suggestions for a project which needs considerable time, thought and effort; it may appeal to those who feel that budding psychologists ought to be encouraged to venture occasionally into the outside world and to bring back material which is not entirely predigested.

The literature on social class being very extensive, references will be confined to works directly relevant to the experiments. For background material the reader may wish to consult the bibliography by MacRae (1953).

SOCIAL CLASS AND PERCEPTION

This experiment was inspired by the work of Asch (1952) on the way impressions of personality come to be formed. He arrived at two broad generalizations: (a) certain characteristics such as the 'warm/cold' dichotomy are central, in the sense that they colour the mode of perceiving all the other (peripheral) characteristics; (b) 'The views we establish of persons are, to a high degree, a function of their group-membership and group position' (op. cit., p. 219). In the present experiment both principles are utilized by making group-membership itself the central characteristic, which is systematically varied.

Preparation. The 'personality descriptions' forming the key stimuli are shown below. In each case the five neutral items are identical, whilst the remaining one constitutes the central characteristic indicative – yet unobtrusively so – of social class.

NEUTRAL ITEMS	VARIANTS OF CENTRAL ITEM
Is forty-three years old	
Has a wife and three children	
Served in the Forces	Owns a large country house
	Owns a semi-detached house
Enjoys pictures	Lives in a council house
Is keen on sport	Lives in a crowded tenement

These 'personality descriptions' are conveniently duplicated on single quarto sheets, which can then be cut into four small slips to be arranged in rotating order, i.e. country, semi, council, tenement, and so on (these abbreviations will be used henceforth in referring to the variants). Next, copies of the Personality Inventory reproduced on p. 158 are needed, also a set of blank quarto sheets and foolscap envelopes.

In preparing the material, an Inventory is folded and put into the envelope; then a 'personality description' slip is folded and added,

but kept near the opening so that it remains easy to reach. When everything is done there should be a set of envelopes so arranged that the different variants recur throughout the pile in the order given above. Experimenters should have a stopwatch, or one with a second hand, and it is advisable to carry spare pencils.

The subjects whose results will be used to illustrate the experiment were members of Workers' Educational Association classes, but it has also been done with university students, and in fact any group with a reasonable degree of literacy is suitable.

Administration. It is preferable, though by no means essential, that subjects should be seated some distance apart. Groups larger than about 30 are difficult to handle. The total number of subjects to be aimed at for obtaining significant results should not be much less than 100. *Please follow ALL instructions EXACTLY.*

Owing to the paramount importance of the mode of administration in this experiment, instructions are given verbatim.*

The object of this short experiment is to determine the extent to which people are capable of sizing up a person from just a few facts about him. You will be given brief notes about a certain individual, and we shall be glad if you will describe in a paragraph or two what kind of person you think he is.

[Distribute envelopes and quarto sheets, one of each per subject. As soon as distribution starts, say:]

Please do not open the envelopes until I tell you to do so. There are two papers in them, one small one ready to be pulled out, and one larger one; *for the first part you are only to take the small one*; leave the larger one inside. The small paper contains the facts about the individual – it is not possible for me to give you any other information apart from what you will see on the paper. There will only be 20 seconds for scrutinizing this paper, then you put it back and write the personality sketch on the blank sheet provided. When you have read the few facts, you will no doubt get some impression of the person, and perhaps form a mental picture of

* The instructions are of course *not* meant to be read, but must be carefully memorized beforehand.

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

him; it is this we should like you to write down, and you are quite free to fill it out with your imagination. Are there any questions?

Now please get ready to pull out the small paper – you have 20 seconds to look at it. Start *now* . . . please put it back into the envelope, and write the personality sketch. You have 10 minutes to do this.

[Give a warning 2 minutes before time is up. Then say:]

Please fold up the sheet; take out the large folded sheet from the envelope, but *do not open it until I say so*; and *put the one you have just completed into the envelope*.

The last task we should like you to do is the completion of a short personality inventory about the same individual. It consists of pairs of statements; please *underline* in the case of each pair the statement which you believe is more likely to apply to the person you have described. If you are not sure, just guess, but please don't leave any out. What we want here are snap judgements, so we shall go through the list together and only about 10 seconds will be available for each choice.

Please open the sheet now, and we begin with No. 1. Underline *either* the statement on the left, *or* the one on the right.

[Continue to call out No. 2, No. 3, No. 4 . . . every ten seconds until No. 20.]

Please make sure now you have not left any out. Kindly fold up the inventory and put it back into the envelope, which should now contain three different bits of paper.

PERSONALITY INVENTORY

Please *underline* in the case of each pair the alternative which in your view is likely to apply to the person you have described. Don't spend much time on any one; if you are uncertain, just guess. BE SURE NOT TO OMIT ANY.

- | | |
|---------------------------------------|--|
| 1. Mainly an optimist | Mainly a pessimist |
| 2. Regards his work lightly | Conscientious in his work |
| 3. Spends much time with his children | Usually leaves them to their own devices |
| 4. Tends to be thrifty | Rather reckless with money |
| 5. Rarely helps with housework | Often helps in the house |

SOCIAL CLASS

6. Lives mainly in the present	Plans for the future
7. Attentive to his wife	Apt to take wife for granted
8. Fairly fond of gambling	Opposed to gambling
9. Self-reliant	Dependent on others
10. Somewhat untidy	Meticulous in his habits
11. Largely self-centred	Great concern for others
12. Active church member	Not bothered about religion
13. Loud and boisterous	Quiet and reserved
14. Shares his wife's interests	They go their own ways
15. Left in politics	Right in politics
16. Slow and deliberate	Quick and impulsive
17. Somewhat ambitious	Has few ambitions
18. Rather patriotic	Not very patriotic
19. On friendly terms with his neighbours	Tends to remain aloof from his neighbours
20. Scrupulously honest	Not averse to petty dishonesty

A few additional comments may be helpful. It is desirable to lay considerable stress on the need to use one's imagination, as some subjects assume that the description refers to a real person and are inhibited by the fear of going wrong; however, it would be a mistake actually to say that the person to be described does not exist. Subjects may often wish to ask for information about this or other points, especially the meaning of words: this has to be avoided at all cost, and one must fend off questions by promising to answer them at the end. The reason will become evident if one considers the hypothetical case of someone being allowed to ask whether the word 'pictures' refers to films or paintings; the chances are that this possible ambiguity would never even have occurred to the majority of the group; but once their attention is drawn to it, responses would thereby become distorted. It is because questions of any kind are precluded as soon as the material is seen, that clarity in giving instructions is vital.

After the results have been collected, the outside of the envelopes, and each of the three papers inside it, should be marked with a code number.

Analysis of the Personality Inventory. The first step is to prepare a summary table along the lines shown below, where L = left and R = right.

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

Item No.	COUNTRY		SEMI		COUNCIL		TENEMT	
	L	R	L	R	L	R	L	R
1.	28	1	26	2	23	6	28	3
2.	5	24	2	26	11	18	13	18
—								
—								
20.	22	7	19	9	8	21	16	15

It will be apparent that the system of rotation produced a sufficiently even distribution of the four 'personality description' variants among the subjects. A broad overall view of the results may now be obtained by computing the percentages of subjects in each category who underlined the statement, say, on the left. In the present example this would look as follows (in rounded percentages):

Item No.	COUNTRY	SEMI	COUNCIL	TENEMT
	%	%	%	%
1.	96	93	79	90
2.	17	7	38	42
—				
—				
—				
20.	76	68	28	52

One can see at once that the first item, which is really a dummy one, fails to reveal any sizeable differences, whilst the opposite is true of the second one. It is, however, necessary to check this impressionistic picture by significance testing. The data for each question may be cast into a 2×4 contingency table, and chi-squared applied.*

A sample table incorporating responses to Item No. 2 is set out below:

	COUNTRY	SEMI	COUNCIL	TENEMT	TOTAL
Observed left	5	2	11	13	31
Observed right	24	26	18	18	86
Left and right	29	28	29	31	117

df = 3; $X^2 = 12.41$; P less than .01

* Where a large number of similar computations are to be carried out by different students, the writer has found it convenient to prepare a computing sheet, which ensures uniformity.

In the experiment whose results are quoted here for purposes of illustration, 13 out of the 20 items yielded differences significant at or beyond the 1 per cent. level. Apart from the overall significance level, it is most instructive to scrutinize the several components of the chi-squared, in order to see which of the four variants made the greatest contribution towards the total.

What to do with the personality sketches. These are valuable even if they are merely read for the insight they provide into the kind of stereotyped images subjects build up after exposure to the stimulus variables. Here is a 'country house' example:

A fairly common type of personality – born into the upper middle class – Eton and Sandhurst or more likely Eton and Oxford or Cambridge. Works in stockbroking, insurance, shipping at or just below partner/director level. Not awfully brainy, but well-connected on an old-boy basis with other 'gentlemen in business'. Played most games when younger – now mainly golf – still aspires to odd game of polo when it can be arranged. Came out of army at end of war as a major, but has dropped the title.

Awfully conscious of what is and is not done. Very fond of his family. Loves his wife in an awkward undemonstrative way. Spoils his daughters but brings up his son very well.

His liking for pictures is the oddity in the pattern. Maybe he wanted to be an adventurous bachelor as well as a family man and this gives him a subconscious outlet.

More honest than most, but doesn't think hard about fundamentals – it's too disturbing.

The family keeps dogs and he likes to show his flowers. A sociable man to know, and he won't let you down.

The sketches often help in the understanding of deviant inventory responses, and the 'perceptual defences' studied by Haire and Grunes (1950) can sometimes be observed, as in the above quotation in connection with 'enjoys pictures'.

Several more systematic approaches can be used. One of these is content analysis, briefly described in the last section of this chapter, whereby the spontaneous stereotypes of the subjects can be extracted.

Another way would be 'blind' sorting according to the variant which has evoked the sketch; before embarking on this all sketches containing references to the housing variable must be eliminated. The extent of agreement among judges can then be worked out.

It will be appreciated that the process of judgement in such a situation involves very complex relationships: the judges' own stereotypes, and what they believe the subjects' stereotypes to have been, all contribute to the outcome. Hence the reports of the judges about how they perceived their task, and the criteria they ostensibly used, should be obtained.

One of the more subtle problems concerns the nature of the scale, and a method of investigating this will be outlined. The first step is to get pairs of judges (or more if there is enough manpower) to carry out the blind sorting task described above *jointly*, in order to lessen the potential effects of individual bias. In practice they always do much better than chance (in the present example 63 per cent. right as against 25 per cent. expected), but the interesting aspect is the pattern of errors. If one assumes that the range from tenement to country house constitutes an equal-interval scale, it is possible to establish the expected frequencies of errors of different types by attaching a probability inversely related to scale distance to each; e.g. tenemt - council = .21, tenemt - country = .07. Below, observed are compared with expected frequencies (rounded to nearest whole number):

<i>Types of Errors</i>	N of OBSERVED <i>Errors</i>	N of EXPECTED <i>Errors</i>	O-E
Tenemt-council	31	16	+15
Council-semi	13	16	-3
Semi-country	17	16	+1
Tenemt-semi	7	11	-4
Council-country	6	11	-5
Tenemt-country	1	5	-4
	75	75	

The two distributions obviously fail to match, and the highly significant difference is in fact largely accounted for by the excess of

tenement - council' errors. It follows that the original assumption is shown to be untenable, the psychological distance between these two being (for the particular subjects concerned) significantly smaller than any of the others. A conclusion of this kind can be checked by reference to the Personality Inventory responses.

Finally, it may be well to point out that the results of this experiment will be to some extent a function of the social class affiliation of the participants.

Suggestions for variations. The first is directly implicit in the last paragraph, or sex-differences may be investigated; for this the sample has to be fairly large, however.

The Personality Inventory can be cut, expanded, or changed by the inclusion of different items. On the other hand one has to be very cautious in altering the basic personality descriptions, because it is difficult to anticipate the effects of any changes. For instance, in one experiment the phrase 'smokes and drinks' was included; this resulted in an image of self-indulgent character which often overshadowed the key socio-economic variable.

If the time available for administration is short, the writing of the personality sketch can be omitted after suitable modification of the instructions; but this is not recommended.

FRIENDSHIP CHOICES AT SCHOOL

The aim of this experiment is to demonstrate the application of sociometric techniques. A few notes on the literature will be given at the outset: Moreno (1934, 1955) - a classical text; Bjerstedt (1956) - authoritative work on classroom investigations, with balanced historical and critical survey; Northway (1952) - elementary introduction intended for teachers; on general methods and statistical analysis: Proctor and Loomis (1951), Lindzey and Borgatta (1954), Criswell (1950), Nehnevajsa (1956).

Classroom studies in Britain have been described in a work edited by Fleming (1951) and more recently a comprehensive survey by Blyth (1960) has appeared, but the enquiry most relevant here is that of Silberman and Spice (1950), as the present experiment owes

much to their mode of approach. They obtained a rating of children's clothing standards and found that 'children of the superior clothing group do discriminate against those of the inferior clothing group in their friendship choices' (p. 57). In so far as clothing standards are presumably associated with socio-economic differences – which were not independently assessed – this indicates that friendship choices were influenced by social class.

It is only fair to warn the reader at once that such social class effects are on the whole not very pronounced among school children in this country, and therefore cannot be readily detected by a small-scale experiment. The main reason for this is that by and large the schools, especially in urban areas, tend to cater for children coming from a fairly homogeneous social milieu. One should not expect social class influences to be clearly manifest unless certain conditions are satisfied: socio-economic differences should be substantial, and not clustered on the margin separating manual from non-manual; a mere sprinkling of children from a contrasting social background is not enough – they must form a substantial subgroup; the schools should either be mixed, or for girls only.

If one has no access to schools which meet these specifications and is anxious to get positive results, which is certainly good for morale, it is inadvisable to focus the exercise on social class. The justification for the inclusion of this experiment is, however, that exactly the same techniques may be employed for a wide variety of other studies, e.g. sex or colour cleavages in friendship choices.

Preparations. The first essential is to seek the permission of the Director of Education to carry out experiments in schools. If parental socio-economic status is to be the key variable, one should avoid any misleading impression that children might be asked personal questions about their parents; if necessary, it should be explained that only the father's occupation is needed for subsequent analysis.

In selecting particular forms, several factors must be considered. The older the children, the greater their awareness of social differences, and the more suitable they are as subjects from this point of view; but in a school with children from a heterogeneous social background there tends to be a creaming off after the age of about nine.

Very young children in the infant school do not yet have a proper understanding of the word 'friend', nor can they write. The best compromise is to take children between the ages of eight and ten.

The children will be asked to send something to their 'best friends', and in order to enhance motivation it is desirable to make the procedure as realistic as possible. Thus, if the experiment takes place near Christmas, then Christmas cards might be sent, or picture postcards at other times of the year.* It is useful, moreover, to symbolize the distinction between best and second-best friend by a corresponding grading in the cards, e.g. one being coloured and the other plain. The total number of cards required, with two choices, is of course twice that of the number of children in the classes to be included in the experiment, plus some ten per cent. extra; this is partly for spoilt cards, and partly for another reason to be explained later. For the 'collection' mock postboxes can be constructed from cardboard.

Administration. At least two and not more than three students should be assigned to any one classroom. One, preferably with some teaching experience, is made responsible for introducing the task, whilst the other(s) assist the children and make sure they follow the instructions. The introduction might run roughly along these lines:

We have come to find out how well children in this school can write (Christmas/post) cards to their friends. Have any of you written cards or letters before? [Pick a particular child, ask for his name and encourage him to tell his story; do the same with one or two others.]

Now here we have two nice cards, one coloured and one plain. We should like you to send this coloured one to your best friend *in this class*, and this other card to another friend *also in this class*. When you get the first card, you put the *full name* of your best friend on the right [indicate], and on the left you write GREETINGS FROM [put on the blackboard]; underneath that you write

* An obvious objection is cost, but Christmas cards can often be purchased cheaply when the season is past, and most large stationers have stocks of old postcards of which they are willing to dispose at a very low price.

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

your own full name, so that they know who has sent it. When you have finished the first card, you put it on your desk and get the second one. On this you put the name of another friend *in this class*, in just the same way. [Repeat instructions; if Christmas cards are used, explanations should of course be suitably adjusted.]

When you have done all that, you will put the cards into the letter-box we have here, and tomorrow/next week we shall come back and deliver them to the friends to whom you have written. [Distribute first coloured card and repeat main directions, emphasizing *best friend in this class*.] Has anyone any questions? Right, start now and write the card to your best friend *in this class*; when you have finished, put the card on your desk. [Wait till all the children have completed the first card, checking meanwhile.] Now you will get the second card, and you write this just the same as before to *another friend in this class*. [Distribute plain cards, supervising completion. At the end, each pupil should have two signed cards on his desk, addressed to two different friends; this should be finally verified whilst going round with the letter-box, immediately before the cards are being dropped in.]

Coding and return of cards. Information as detailed as possible will have been obtained about parental occupations, and this will have been entered into the class list, on which the children absent during the experiment should also be marked. From this a new list is prepared, where the children are divided according to sex and parental occupation; names *within* each subgroup should be arranged in alphabetical order, as shown below for boys:

Non-manual(X)		BOYS (B)	
Name	Code	Manual(Y)	Code
Adams, John	BX 1	Armstrong, Fred	BY 1
Clark, Robert	BX 2	Burton, William	BY 2
.....
Unwin, Graham	BX n	Taylor, Jack	BY n

The next step is to assign a code, as illustrated, and on the basis of this coded list all the information is then transferred from the postcards onto filing cards. For instance, if Adams addressed his

first card to Clark and his second to Armstrong, the filing card would, look as follows:

<i>Sender</i>	<i>1st choice</i>	<i>2nd choice</i>
BX 1	BX 2	BY 1

The completed set of filing cards is sorted by sender into four packs (BX, BY, GX, GY), the cards *within* each pack being arranged in numerical order (e.g. BX 1 to BX n). Once this has been done, and checked, the original cards are no longer needed and are almost ready for 'delivery'. Before that, however, another job needs to be done: when comparing the postcards with the original class list during checking, find out which children are isolates, i.e. those to whom no one sent anything. It is good practice to 'forge' some cards addressed to them, omitting any signature, so as to avoid causing any heartburning.

Construction of sociomatrix. This is most conveniently done on a sheet of graph paper, and the general layout can be seen in Fig. 4, where data from an experiment in Primary V of a village school are plotted. The full horizontal and vertical lines mark off the sexes, the broken ones socio-economic status within each sex category.

Entries into the matrix are made directly from the pre-arranged filing cards; thus BX 1 had as his first choice BX 2 (symbolized as 10) and for his second choice BX 7 (symbolized as 01); and so on for all the rest. Note that where the number of available choices $d = 2$, all right-hand marginal totals must add up to this. Choices directed at individuals outside the class should be shown beyond the right-hand margin of the matrix, as was done in the present instance with the first choices of GY 3 and GY 7. The number of first and second choices received by each subject is listed at the bottom; the total, including outside choices, will of course equal dN , where N is the population of subjects.

The diagonal line permits the rapid identification of mutual choices, which are always symmetrically situated on either side of it; e.g. BX 1 chose BX 2 and vice versa. All mutual choices in Fig. 4 are circled.*

* No distinction has been made between pairs of first, second, and mixed choices, but they can be differentiated if desired.

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

Quantitative analysis. A great deal can be learnt from the mere inspection of a rationally designed sociomatrix. Among the sixteen fields of the present matrix, seven of the cross-sex ones are com-

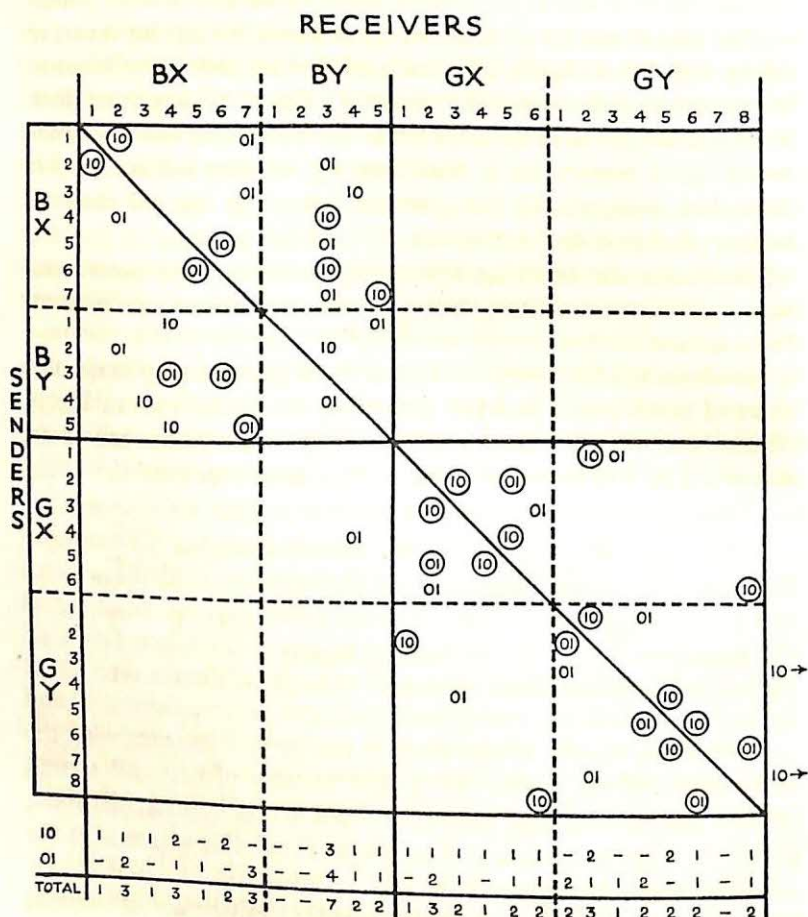


FIG. 4.

pletely empty, and the eighth contains only a single entry. There is thus such a high degree of sex-cleavage that significance testing is hardly necessary.

SOCIAL CLASS

The chief concern in this experiment is to investigate the influence, if any, of parents' socio-economic status. One way of approaching this would be to ask whether on the average 'manual' or 'non-manual' children tend to be more popular, i.e. highly chosen. Numbers are far too small to permit any significance testing; however, it may be noted that choices received in each of the four sub-divisions cluster very closely around the expected value of 2, suggesting that the answer would be in the negative as far as this population is concerned. This impression is confirmed by an examination of the 'stars' and 'isolates': all three isolates belong to the non-manual category, but so does the only star.

The other, quite distinct, question may be phrased as follows: was there a significant tendency to choose one's friends among children with a similar socio-economic background? Scrutiny of the distribution of choices in the sociomatrix suggests that boys and girls should be tested separately. The actual procedure is simply to take the top left and bottom right-hand quarters of the sociomatrix and treat each as a 2×2 chi-squared table, in the manner set out:

		BOYS				GIRLS	
		<i>Receivers</i>				<i>Receivers</i>	
		BX	BY			GX	GY
<i>Senders</i>	BX	7	7	<i>Senders</i>	GX	8	3
	BY	7	3		GY	3	11

There is clearly no association as regards the boys' choices, and significance testing would be redundant; but with the girls chi-squared (corrected for continuity) is significant at the 5 per cent. level.*

It is further possible to use the data for testing the null hypothesis as regards a sex-difference in the tendency to make in-group choices related to socio-economic status; this turns out to be significant at the 2 per cent. level.†

* For a more exact test, cf. p. 178.

† In these computations the two outside choices were ignored. If they are included, and it is assumed that they go counter to the

The material can also serve to demonstrate the use and limitations of a probability model in sociometric work, and the example given here relates to the frequency of mutual choices. At the outset some of the fundamental notions will have to be introduced.

(a) The number of ways in which two elements can be selected out of a class of N elements, where N is the number of members in the group, is C_2^N ; in other words, this represents the total number of possible pairs:

$$C_2^N = \frac{N(N-1)}{2}$$

(b) The probability that any particular individual will choose any other particular individual is p ; the probability that this will not happen is q ; the number of choices permitted to each individual is d :

$$p = \frac{d}{N-1} \quad q = 1 - p = 1 - \frac{d}{N-1}$$

(c) The expected number of unreciprocated choices in a population is $E(Ur)$:

$$E(Ur) = pq \cdot N(N-1)$$

(d) The expected number of mutual pairs is $E(Mp)$:

$$E(Mp) = p^2 \cdot C_2^N = p^2 \cdot \frac{N(N-1)}{2}$$

In the present case these formulae may be entered with $N = 26$ and $d = 2$. Then $E(Ur) = 47.84$; $E(Mp) = 2.08$; note as a check that:

$$\frac{E(Ur)}{2} + E(Mp) = 23.92 + 2.08 = N$$

Thus the number of mutual pairs expected on the assumption of randomness is 2.08, whilst the actual number of mutual pairs observed (m) is seen from Fig. 4 to be 14. A formula to test the

tendency observed, the first difference would cease to be significant, and the second would be reduced to the 5 per cent. level. This underlines the desirability of avoiding outside choices, especially when the samples are small.

significance of the difference directly is given by Proctor and Loomis (1951, p. 578):

$$X^2 = \frac{(m - C_2^N \cdot p^2)^2}{C_2^N \cdot p^2 \cdot (1 - p^2)} \quad (\text{with one degree of freedom})$$

hence
$$X^2 = \frac{(14 - 2 \cdot 08)^2}{2 \cdot 08 \cdot 0 \cdot 9936} = 68 \cdot 74$$

Such a fantastically high figure for chi-squared is unfortunately no cause for rejoicing; on the contrary, it indicates that a test of this kind is really superfluous with anything that can properly be called a group. More useful would be some measure of the probability of obtaining by chance a level of mutuality as large or larger than the one observed; but this appears to be an exceedingly complex problem, which has so far not been solved.*

This example should also provide a caution against any attempt to regard the frequency of mutual choices as a measure of group-cohesion or group-integration: in spite of the highly significant number of mutual choices, the present class-group was internally divided by an almost complete sex cleavage.

Suggestions for variations. For reasons of space these cannot be considered in detail, but Bjerstedt (1956) will be found a fruitful source of ideas. It might perhaps be added that with older children the letter-writing situation is apt to be somewhat artificial, and it is easier to arrange a mock election where the post-box is replaced by a ballot box.

CHILD-REARING PRACTICES

This is envisaged as a co-operative task, of which the initial planning forms an integral part. Hence it is proposed to offer merely a series of hints about lines of approach and some of the snags liable to be encountered.

Preliminary work. Those participating should make themselves acquainted with the literature about social class differences in this sphere. The bulk of it is American, and the somewhat contradictory

* Cf. Katz, Tagiuri, and Wilson (1958).

findings are summarized and evaluated by Bronfenbrenner (1959). In Britain there are studies by Spinley (1953) and Kerr (1958), mainly concerned with sections of the working class; an article by Shapiro (1952) is also relevant, and much can be culled from a publication by UNESCO (1949).

A decision has to be made about the scope of the project. It may be of a general exploratory nature, covering a fairly wide range of topics and being primarily descriptive in its purpose; alternatively, a number of specific hypotheses may be formulated, and the project focused on their testing.

Each participant may be asked to draw up his own interview schedule, which is then circulated to all others. An agreed version can be arrived at by joint discussion, a group task which is often most instructive.

Choice of informants and role of interviewer. Both these issues have to be cleared up before a realistic and workable schedule can be prepared. On the first point the standard research procedure, namely proper sampling, must be rejected, because its difficulties in this sphere are such that the project would become hopelessly unwieldy. Instead, it is suggested that each participant should himself contact, say, two middle-class and two working-class informants from within his social orbit.* The most effective criterion of social class is a simple 'manual versus non-manual job' dichotomy; and as students are more apt to pick on members of the 'aspiring' working-class, it is desirable to specify that the wife should also have been in a manual occupation before marriage.

Another general requirement is, of course, that there should be pre-adolescent children in the family. These specifications represent a compromise: they leave considerable leeway for variations of many kinds, but it must be remembered that the more details are laid

* Experience has shown that one can normally expect this from psychology students. Problems sometimes arise with those living away from home, who may be at a loss to find suitable working-class informants. In that case suitable arrangements can usually be made by putting them in touch with a university settlement or similar institution.

down, the harder it becomes to find informants who conform to all of them. As they stand, sufficient control is ensured for the reasonable expectation that some general patterns may emerge.

It is taken for granted that the person to be interviewed will be the mother; there is, however, the possibility that the father may also be present, and means of handling this situation have to be considered.

Some thought must always be devoted to the question of the interviewers' role, on which explicit agreement has to be reached. Otherwise some might adopt the questionable role of the 'friend' who tries to obtain information in a casual-seeming manner in the course of ordinary social relations, whilst others might be tempted to don the formal mantle of 'researcher'. In fact the plain and straightforward solution is probably also the most effective: the role of a student who has been given an assignment and asks, as a favour, the help of the informant in fulfilling it.

Similar considerations apply to the mode of introducing the topic of the interview. Details will of course vary according to the particular theme planned, but there should be uniformity in the preliminary explanations. Mothers are generally interested in problems of child-rearing, and it is usually not very difficult to secure their co-operation. References to social class are, however, best avoided. It may also be mentioned here that informants should be given an assurance of anonymity, though in a casual manner. Inexperienced interviewers sometimes stress this unduly, and may thereby actually suggest to the informants that the enquiry would be of a personal and intimate nature.

Drafting an interview schedule. Whether the enquiry is of the general exploratory kind, or whether specific hypotheses are framed, it is advisable in a project of this type to leave things like breast feeding and toilet training alone. Even after such a self-denying ordinance, one of the major pitfalls is usually excessive length: accumulating data is relatively easy, processing them laborious and time-consuming. From this point of view the limited hypothesis approach is perhaps preferable, as it reduces the temptation to wander into too many different directions. An example of such a hypothesis would

be that mothers at different socio-economic levels differ in the sources from which they acquire their ideas about how children should be brought up. A number of questions relating to this are then devised, to be preceded in the actual interview by some other general and neutral ones, which gradually lead up to the key theme during the warming-up stage.

A few brief suggestions will also be made about the kind of topic that might be included in a broad exploratory interview:

Expected ages of speaking/reading simple sentences;

Bed-times and amount of pocket money at various age-levels (e.g. 5, 10, 15);

Ages at which children did/should:

wash and dress themselves

travel on buses alone

go out to pictures alone in the evening

go out with boy/girl friend

Proposed school-leaving age;

Kinds of misbehaviour parents find most annoying;

What sorts of punishments used at different ages by whom?

What sort of help expected in and around the house at different ages?

Kinds of jobs envisaged/desired for offspring.

It will be noted that the first few are in such a form as to yield quantitative responses directly; the analysis is considerably shortened if a reasonable proportion of this type are included.

Needless to say, the above is only the sort of raw material about which the actual questions can be constructed. Great care must be taken in the phrasing, as difficulties may arise from the heterogeneous character of the informants' families. Thus one has to decide which questions ought to be asked separately about boys and girls; or again, which ought to be about actual practices, and which about general norms. Depending on the ages and sexes of the children in each family, the same question will unavoidably sometimes have to be factual for some informants and hypothetical for others.

Questions should always be formulated with a certain amount of

tact and cunning, especially when one answer is clearly more socially approved than another; thus do *not* ask: 'Do you ever help your children with their homework?' but ask instead: 'Do you find time to help, etc.?' In this kind of work interviewers should aim at more than a 'Yes' or 'No' answer, and the questions (other than purely quantitative ones) should be so designed as to stimulate a certain flow of talk. Under such circumstances it is difficult to record everything verbatim, but the attempt should be made; for answers often only become meaningful within the context of ideas from which they spring. For the same reason the interviewing plan should not be unduly rigid, with an unalterably fixed sequence of questions; it is best to aim at an easy, semi-conversational style, which will not inhibit informants.

Details about social background should be obtained at the end of the interview, and recorded on a separate sheet that can later be detached. As a minimum the following might be suggested: occupation and education of both father and mother; ages, sexes, and schooling of children. If the interviewers are reasonably skilled, the occupational level of grandparents might also be included; this can bring out cases of social mobility, which are often particularly interesting, but some tact is needed because socially mobile people are sometimes reluctant to divulge the fact.

Once the interview schedule has been hammered out, it must be put to the test in some trial interviews, which will at the same time be useful practice runs for the interviewers.

Analysis of responses. Completed interview schedules should be given code numbers on the front page and on the sheet containing social background details, which is then detached. This needs to be done in order to exclude the possibility of bias in the categorization of qualitative material.

All numerical responses can be directly subjected to statistical treatment, and the non-parametric methods described by Siegel (1956) are admirably suited for this kind of material. Actual responses will be couched in terms of various units such as hours, years, or shillings, and these should be transformed into decimals; where, as often happens, a range is given, one simply takes a median value; e.g. 'between seven and half-past' becomes 7.25.

Qualitative responses, which may be lengthy and multi-dimensional, are more difficult to handle. The best procedure to adopt is one outlined by Raven (1950). This involves copying each answer (or with long ones a condensed version of it) on a separate card, marked with the code number. Cards with similar content are then placed in the same column, and whenever a different kind of answer is encountered a new column is started. If there is any doubt in the case of condensed version, reference back to the full answer may provide the solution. In this way a series of 'natural' categories are allowed to emerge. Once this has been done, the resulting categories can again be treated statistically. However, as has been repeatedly emphasized, it is a great mistake to become blinded by numbers, overlooking the fact that much can be learnt from the manner and content of responses, even if the differences fail to reach the conventional level of significance.

A few concrete illustrations will now be given. First, a question about children's bed-times; this was chosen as a check on the work, because it is known from the large-scale survey of Himmelweit and others (1958) that a class difference does exist in this respect.

'What would you say is the time children of 5 ought to go to bed? And when they are 10 years old? And at 15?'

The responses relating to five years, transformed into decimals and with ranges eliminated, were as follows:

Non-manual: 6.0, 6.0, 6.5, 6.5, 6.5, 7.0, 7.0, 7.5, 7.5;
Manual: 6.0, 7.0, 7.0, 7.5, 7.5, 8.0, 8.0, 8.0, 8.25, 8.5, 8.5;

The Mann-Whitney U test was applied to this and the other two series, with the result that the differences were found to be significant for ages five and ten, but not fifteen.

The next question is an example of a less straightforward type:

'Should one leave a child free choice with whom to play, or is it necessary to discourage him from playing with *some* children?'

Use of the sorting method yielded at first a simple tripartite division:

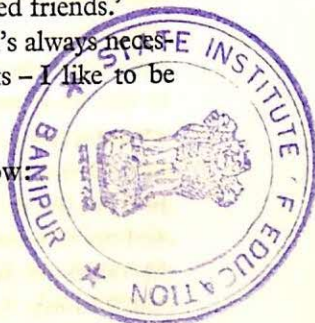
A Those favouring unqualified freedom: 'My children pick their own friends; they can bring anyone home they want.'

SOCIAL CLASS

- B Those favouring freedom in principle, yet having some reservations: 'I leave the choice of friends to them, but would interfere if they brought home very dirty or foul-mouthed friends.'
- C Those favouring definite restrictions: 'I think it's always necessary to discourage them from certain elements - I like to be particular about playmates.'

The distribution of these categories is shown below

	A	B	C
<i>Non-manual</i>	4	3	7
<i>Manual</i>	3	7	2



The trend seems to be in the expected direction, i.e. a greater frequency of definite restrictions among non-manual mothers, though the numbers are too small for significance testing. One might leave it at that, missing thereby some valuable insights. If one does go on, the next operation is to re-arrange the cards *within* the two categories B and C so as to find out how the reasons compare. In the present case some of the contrasts were illuminating; for example, take two informants favouring definite restrictions: the non-manual mother said she 'wouldn't like her children to get a horrible accent', whilst the manual mother was anxious to keep her offspring away from some older children who had been in constant trouble with the police.

This will help to understand why interviewers must both encourage full and free answers, and record them in sufficient detail, if the material is to be used to elicit the subtler aspects just touched upon.

Lastly, an example of a question designed as part of a test of several hypotheses about sources of information on child-rearing:

'Where would you say you have got most of your ideas from about the way children should be brought up?'

Here the sorting method is employed again, but with the difference that the categories are now pre-determined by the hypothesis; this does not mean, of course, that one should not keep on the look-out for the unexpected. In the present case it had been postulated that non-manual mothers would be more likely than manual to resort for

guidance to printed sources, i.e. books and magazines. When the responses were sorted accordingly, the following pattern resulted:

	BOOKS AND MAGAZINES	
	<i>Mentioned</i>	<i>Not mentioned</i>
Non-manual	9	5
Manual	2	10

Tables of this kind, too small for chi-squared, may often arise and the Fisher-Yates test of significance in 2×2 contingency tables is the appropriate one to use. Where marginal frequencies do not exceed 20 this can be done with a minimum of effort by looking up the tables prepared by Finney (1948) and Latscha (1953). In this case $P = .025$ (one-tailed test), and the hypothesis is thus supported.

Several other variables such as clinics, doctors, teachers, were considered according to a previous plan; no hypothesis had been formulated regarding the influence of the informants' own mother which, it had been tacitly assumed, would be universally acknowledged in all social classes. In the course of scrutinizing the responses, however, it appeared that this was mistaken, as appears in the table below:

	OWN MOTHER	
	<i>Not mentioned</i>	<i>Mentioned</i>
Non-manual	13	1
Manual	6	6

This difference also turned out to be significant at $P = .05$ (two-tailed test), and is perhaps more interesting than the others. It is the excitement of such small discoveries, however insignificant they may appear *sub specie aeternitatis*, which makes projects rewarding and compensates for the time and trouble involved.

REFERENCES

- ASCH, S. E. (1952). *Social Psychology*. New York: Prentice Hall.
 BJERSTEDT, A. (1956). *Interpretations of Sociometric Choice Status*.
 Lund: Gleerup.
 BLYTH, W. A. L. (1960). 'The sociometric study of children's groups in English schools'. *Brit. J. Educ. Studies*, 8, 127-47.

- BRONFENBRENNER, U. (1959). 'Socialization and social class through time and space'. In Maccoby, E. E., Newcomb, T. M., and Hartley, E. L. (eds.): *Readings in Social Psychology*. London: Methuen.
- CRISWELL, J. H. (1950). 'Notes on the constant frame of reference problem'. *Sociometry*, 13, 93-107.
- FINNEY, D. J. (1948). New Statistical Table No. VII from *Biometrika*.
- FLEMING, C. M. (1951). *Studies in the Social Psychology of Adolescence*. London: Kegan Paul.
- HAIRE, M., and GRUNES, W. F. (1950). 'Perceptual defenses: processes protecting an original perception of another personality'. *Human Relations*, 3, 403-12.
- KATZ, L., TAGIURI, R., and WILSON, T. R. (1958). 'A note on estimating the statistical significance of mutuality'. *J. Gen. Psychol.*, 58, 97-103.
- KERR, M. (1958). *The People of Ship Street*. London: Kegan Paul.
- LATSCHMA, R. (1953). New Statistical Table No. XVII from *Biometrika* (available from Biometrika Office, University College, London).
- LINDZEY, G., and BORGATTA, E. F. (1954). 'Sociometric measurement'. In Lindzey, G. (ed.): *Handbook of Social Psychology*, Vol. I. Cambridge, Mass.: Addison Wesley.
- MACRAE, D. G. (1953). 'Social stratification', *Current Sociology*, Vol. II, No. 1. Paris: UNESCO.
- MORENO, J. L. (1955). *Who Shall Survive?* 3rd Ed. New York: Beacon House.
- NEHNEVAJSA, J. (1956). 'Probability in sociometric analysis'. *Sociometry*, 18, 422-32.
- NORTHWAY, M. L. (1952). *A Primer of Sociometry*. Toronto: Univ. of Toronto Press.
- PROCTOR, C. H., and LOOMIS, C. P. (1951). 'Analysis of sociometric data'. In Jahoda, M., Deutsch, M., and Cook, S. W.: *Research Methods in Social Relations*. New York: Dryden Press.
- RAVEN, J. C. (1950). 'The comparative assessment of personality'. *Brit. J. Psychol.*, 40, 115-23.

SOCIAL PSYCHOLOGY THROUGH EXPERIMENT

- SHAPIRO, M. B. (1952). 'Some correlates of opinion on the upbringing of children'. *Brit. J. Psychol.*, **43**, 141-9.
- SIEGEL, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw Hill.
- SILBERMAN, L., and SPICE, B. (1950). *Colour and Class in Six Liverpool Schools*. Liverpool: The University Press.
- SPINLEY, B. M. (1953). *The Deprived and the Privileged*. London: Kegan Paul.
- UNESCO (1949). *The Influence of Home and Community*. Paris.

Chapter 7

SMALL-SCALE SOCIAL SURVEYS

J. F. MORRIS

University of Manchester

Most of the chapters in this book are concerned with experiments that can be planned and completed in one meeting of a practical class in social psychology. Social surveys, however small in scale, can hardly be reduced to such a narrow compass if they are to be of any value, either as a set of findings or an educational experience. But it is possible to start and finish a survey within a month, using two afternoon meetings and a certain amount of additional time, squeezed into the usual gaps in an undergraduate timetable. Such a survey is illustrated here in some detail. But first it may be useful to say something about social surveys in general terms – their purposes, their history and some of the commoner ways in which they are used and misused.

In essence, surveys are quantitative descriptions of certain aspects of a group of people (or, if we wish to avoid the implication of social organization and possible face-to-face contact, a set of people or a population). Such descriptions stand or fall on two things – their accuracy and their worthwhileness. It is easier to make a survey reasonably accurate than to be sure that it is worth doing. A sense of the theoretical or practical value of survey findings is not easy to teach or to learn. It is the quality that Polanyi (1958) calls ‘connoisseurship’: best developed by doing research and comparing the results with the original conception of the research programme, and by studying the work of skilled research workers, who are masters of their field.

The purposes of social surveys

What are we trying to achieve by a social survey? In general, we are trying to make precise statements about the distribution of opinions, attitudes or activities in a particular set of people in whom we are interested. Our description may be merely superficial and enumerative, failing to distinguish between deep-rooted tendencies and ephemeral responses. Or it may be explanatory, giving us an insight into the interconnectedness of social life. Such surveys require greater skill and insight.

It is easy to depreciate the first kind of survey, but it may serve a useful purpose, especially in the short-run. In market research, for example, a maker of (say) electric shavers wants to know how often men shave, and whether women prefer shavers to depilatories. He wants to know the market for each type of shaving instrument, from the cut-throat razor to the electric shaver. But even here he will be glad if enumeration can be linked with broad trends or to other characteristics of the population – the incomes, ages, occupations and places of residence of the people studied. Certain relationships may then be noted and found to be useful in planning advertising campaigns or in re-designing the product. In addition to these ‘hard facts’ about shaving, it may well be found useful to gain information about people’s attitudes to shaving, particularly those attitudes concerned with maintaining existing habits.

In politics, voting preferences can be linked with information about people’s social background (see, for example, a voting survey of a Derbyshire town by a Manchester University research team – Birch (1959)). The results can be of considerable value to those organizing electoral campaigns, in helping them to distinguish the strategic points at which to direct their efforts. Even if the survey information is based on rather rapidly shifting opinions and attitudes, it can still be useful in a single campaign, though an analysis of trends, combined with last-minute polling, is necessary if high predictive accuracy is to be achieved.

These practical applications of social surveys have increased very rapidly in the last twenty-five years. There is every likelihood that

they will increase still further, particularly in fields relatively neglected at present, such as providing local authority officials and representatives with up-to-date information about ratepayers' views on plans for local development, or measuring community changes in family life (Mogey, 1956; Kuper, 1953; Young and Willmott, 1957).

There is still a tendency for the communication of survey findings to be restricted to the small groups of interested people, such as administrators and business men, who commission the surveys, or to the professional research workers who carry them out. Thus the results do not reach the most important people of all – those who gave the information on which the survey findings were based. Even putting on one side the ethical questions which this practice raises, there still seems to be a considerable loss of possible effectiveness when survey findings are restricted in this way. A survey gives an opportunity of 'feeding back' information into a social group which is often so large, so complicated in organization and so rapidly changing that its members need a continual flow of accurate information in order to adapt themselves effectively to the new situations that are continually evolving. Surveys are not by any means the sole source of such information, but they can powerfully supplement, and often correct, the informal news and views that circulate through the group. By so doing, survey findings increase the possibility of rational behaviour. A conspicuous example of the use of commercial survey data for enlarging the stock of public information is found in the work of Mark Abrams (for example, *The Teen-age Consumer*, 1959).

An historical note

Social surveys in the form of census-taking are very old indeed. Most writers on the history of surveys begin with a few scattered references to Herodotus and the Domesday Book and then get into their stride with John Howard's eighteenth-century studies of prison life, Le Play's nineteenth-century studies of the impact of industrialism on the family life of French working people, and the massive

surveys of urban poverty by Mayhew, Booth, and Rowntree in Britain and Jane Addams, Robert Woods, and Florence Kelley in the United States.

Good brief accounts of this work can be found in Young (1949), Abrams (1951), Caradog Jones (1949), and Moser (1958). Our concern here is not to give an even more condensed survey of surveys than they do, but to note that there seem to have been four themes in the development of surveys. These do not, of course, constitute neat and clearly demarcated stages, but broad themes following and then overlapping one another. The first is the census type of survey, usually carried out by those in power as a form of stocktaking. The Domesday Book and the much earlier survey of the population and wealth of Egypt three millennia before Christ (mentioned by Herodotus) are examples of this kind. The second is the study of the social problems of hunger, crime, and bad housing associated with the uncontrolled growth of towns since the eighteenth century. The third is the community and regional survey movement, which has studied a wide range of social life within towns and regions, emphasizing particularly the links between family, work, and leisure. The fourth theme is the study of socially important choices made by individuals—in voting, and purchasing, for example.

It has already been noted that these themes overlies rather than replace one another. At present, all four themes are to be found in social surveys. It is difficult to generalize, but the attitudes and values associated with them are somewhat different. The first type of survey began as an instrument of central control, often a very autocratic and coercive control. The census no longer has quite these implications, and has indeed become an essential part of our factual knowledge of modern society, without which other surveys would be greatly weakened. But the census is still rather remote from the everyday concerns of ordinary men and women. It is in rather strong contrast, in this respect, to the poverty and crime surveys which were begun and continued with social reform in mind. Often the people who gave the information on which these surveys were based never read the results, nor were they always able to understand the precise aims of the investigation. But they knew that the interviewers were asking

them about matters of great everyday importance – their wages, their conditions of work and housing, and their food. If we consider only such matters as their understanding of the planning and execution of the survey and the effective use of the findings, their collaboration was slight. But they were deeply involved in the problems studied by the survey, and, in addition, the research workers for their part were deeply concerned that the survey should help these people (Simey and Simey, 1961).

The third theme in social surveys – community and regional studies – is also associated with social reform, through re-housing projects, and the improvement of education and other public services. These surveys are concerned with the welfare of the people whose lives are described. It is when we come to the fourth theme that we begin to note different attitudes and values. Studies of voting and purchasing can be made from motives of detached curiosity, or to increase the profits of a particular firm, or to strengthen the position of a political party. Many people fear investigations, particularly privately-commissioned ones, because they fear that the knowledge so gained will be used to manipulate the public in the interests of powerful minorities. At this point, a brief discussion of possible misuses of social surveys seems to be appropriate, since the survey as such is an instrument capable of use and misuse.

The misuse of social surveys

There are three main types of misuse of social surveys: technical errors or faulty assumptions, which affect accuracy; ethical abuses, which violate the rights of those involved in a survey; and errors of judgement, of the type which make a survey a waste of time to everyone involved.

Technical errors are the least controversial, and include biased samples (that is, samples which are not randomly selected), over-generalization of findings to populations not represented by the original sample, unwarrantable inference from inadequate data, and lack of statistical checks on data gained from the survey. An excellent example of a survey critique on technical grounds is to be found in

Sheatsley and Hyman's comments on the Kinsey Reports, in Geddes (1954).

Ethical abuses are controversial in two ways. It is not easy to agree whether a specific practice constitutes an ethical abuse. Again, it is difficult to decide what to do about ethical abuses, when there is a conflict between short-term and long-term aims (as with making confidential or disturbing medical information available to a wide public in a health-education programme) or when there is conflict between technical and ethical requirements. More clear-cut abuses involve such practices as the deliberate withholding of survey findings from the public so that they can be used for commercial or political manipulation. Some of the motivational research surveys seem to come into this category. Much is said by their practitioners concerning the quasi-medical secrecy of the interviews, but the fact remains that the advertisers and commercial interests paying for the research are given the confidential findings and the people whose conscious and unconscious motives form the content of the report are told nothing. Some technical misuse will amount to ethical abuse if it takes the form of deliberate falsification, such as distortion of results, the use of loaded or leading questions, or the suppression of contrary findings. In other cases there may be no technical misuse, but an invasion of privacy in order to secure relevant material; such as the use of hidden tape-recorders, tiny cameras or a concealed one-way viewing screen. These practices are not only ethically repugnant but inexpedient, since they discredit the survey as a technique of research and impair the foundation of trust and goodwill on which social research ultimately rests.

The last type of misuse, triviality and time-wasting, is also controversial. It must never be forgotten that there is always room for disagreement about the worthwhileness of a specific investigation. Moser (1958), for example, quotes a satirical fourth leader in *The Times* on the subject of social surveys.

It has been ascertained and not a moment too soon [the leader-writer states] that out of 1,200 Somersetshire children between the ages of three and fourteen only two-thirds were present when their

shoes were bought. This pregnant intelligence comes as a reminder that almost all over the world, almost all the time, research is going on into the habits and opinions of *homo sapiens* . . . It is rather sad how seldom we are able to take very much interest in the results of these meticulous enquiries.

If the survey in question did no more than assert the proportion of children whose shoes were not fitted in a shoe-shop, then the charge of triviality, or at least incompleteness, seems warranted. But if it formed part of an investigation into the conditions under which children develop foot defects, then it is both interesting and important to know that as many as a third of a sample group of children may be wearing shoes that are ill-fitting from the outset. In fact, a later survey has shown that two-thirds of the children in another sample suffered from crooked toes and other foot deformities. The effects of such defects on general health and well-being can be considerable.

Whether or not a proposed survey is to be judged 'trivial' depends on its relevance or relation to other social facts. Disconnected facts lead us to comment 'So what?' It is the demonstration of a relationship between conditions and consequences, and an indication of the significance of the consequences to human well-being or to some other matter of value, that redeem a social investigation from triviality.

The testing of hypotheses in surveys

As Moser (1958) reminds us, 'it is altogether false to imagine that research divides neatly into two categories – experiments and others – and that only the former can lead to valid causative inferences. Rather should one view researches as being ranged along a scale, with the most completely controlled experiment at one end and uncontrolled observation at the other.' If one thinks of surveys as being mainly of the enumerative kind, there seem to be few hypotheses that they can test, if one takes the term 'hypothesis' to refer to a stated relationship between variables. But a glance at recent surveys will show that they have frequently been concerned with testing hypotheses. Here are a few examples. Willmott and Young (1960) are interested in comparing middle-class family life with working-class

family life. They set up a number of hypotheses, such as the greater importance of the wife's mother in working-class families, and the greater membership of voluntary organizations in middle-class families, and test them by comparing sample data from Woodford, Essex, and Bethnal Green, London. With much more rigorous control, Himmelweit, Oppenheim, and Vince (1958) examine the effects of television-viewing in selected age-groups of children in a number of towns and cities in Britain. By the use of matched pairs (viewers/non-viewers) they are able to test the effects of viewing on school work, aggressive behaviour, leisure activities and other variables. By a rather different process of statistical matching, Belson (1960) examines the effects of television-viewing on such aspects of family life as visiting other families, conversation, and hobbies.

For descriptive hypotheses asserting a reliable association between variables, surveys are well adapted. It is when causal inferences come to be made that surveys compare poorly with laboratory experiments, because the possibility of manipulation is limited in a survey. The result is that critics of surveys can always point to the existence of uncontrolled variables that cannot be assumed to randomize out.

We come now to a small-scale survey that can be carried out by two or more students in a short time. Unlike the experiments detailed elsewhere in this book, the survey deals with people in the ordinary course of social life, not in a laboratory situation. Thus, a certain amount of variability in the procedures used is desirable, in order to allow for individual differences and local circumstances. For example, the interview schedule which is the main survey technique used here, is not elaborated in detail. And the survey report, while suggested in outline, is not rigidly prescribed. The survey to be described is not concerned with simple enumeration, but with examining relationships between certain variables.

A SMALL-SCALE SURVEY

There are ten more or less distinct phases in a social survey: (1) The general problem to be studied. (2) A review of previous work on the problem, which permits it to be stated more definitely in (3) Questions

or hypotheses guiding the survey. (4) The selection of suitable techniques of investigation. (5) The selection of the people to be surveyed. These will usually be a *sample* of a larger population. (6) Testing of the survey techniques in a 'pilot survey' or trial run. (7) The collection of information. (8) The preparation of this information so that the initial questions may be answered or the hypotheses tested. (9) The interpretation of the survey findings. (10) The writing of the survey report.

Each of these phases will now be discussed.

The general problem to be studied. There are many problems that could be studied in a student survey. The university itself is a good starting point. What, for example, is the relationship between student participation in university life and academic achievement? This will be the theme of the present survey. We cannot hope to make a comprehensive investigation into this topic in the time available, but it is possible to examine some limited aspects.

A brief consideration of the terms 'participation' and 'achievement' suggests some of the possibilities – membership of student societies, attendance at lectures outside one's field of study, frequent use of the Student Union, the range of subjects being studied by friends and acquaintances, the degree of interest in student activities, the frequency and diversity of contacts with university teachers, examination grades, the proportion of passes to failures in terminal examinations, and the type of university course (Honours, General degree, etc.). At this point it is useful to consider previous work on the subject.

A review of relevant work. The purpose of this phase of the survey is to see what previous research workers have found out about academic achievement and participation in university life. For example, Himmelweit and Summerfield (1951) and Malleson (1958) suggest that achievement and participation are positively associated. The best way of making a review is to refer to the annual volumes of *Psychological Abstracts* (under the index items of 'Student', 'College Student', and 'Achievement'). Appropriate articles should then be briefly summarized, and a short connected outline should be written as a guide to the next phase of the survey.

The questions or hypotheses guiding the survey. After examining previous work, one might discover that the problem originally envisaged has been amply studied or, more probably, that certain aspects of the problem appear particularly worth following up. Here, four hypotheses are suggested:

(1) Academic achievement is positively associated with participation in the social life of the university, especially the organized aspects.

(2) Academic achievement is positively associated with achievement in other student activities.

(3) Academic achievement is positively associated with an active interest in planning the use of one's time.

(4) Academic achievement will be positively associated with knowledge of university social activities (student societies, union activities, student publications) and aspects of university life outside one's studies.

The selection and development of research techniques. In this phase of a survey, one must decide what methods are best adapted to the questions or hypotheses arrived at in the previous phase. Here is a list of the most frequently used techniques, with some of their advantages and disadvantages.

The questionnaire. The advantages of drawing up a list of questions to which individuals give written answers are: (i) the questions are in standardized form, so that answers can more readily be compared. (ii) those replying to the questions (the respondents) have time to consider their answers before they return the questionnaires. (iii) distributing and collecting questionnaires takes less time than interviewing. (iv) errors of transcription can be avoided, because respondents will have written their own replies. The disadvantages are: (a) it is not possible, without a supplementary questionnaire, to follow up interesting replies or to correct replies that appear to be based on a misunderstanding of the questions. (b) the percentage return of questionnaires is likely to be lower than with interviewing, because people have to overcome their resistance to sitting down to complete the questionnaire (30 to 40 per cent. returns are average for postal questionnaires – short questionnaires, stamped and addressed

envelopes, and, most important, questions that seem important and interesting to the respondents, are factors in gaining a good return). (c) even with a student population composed of people more literate than most, answers to questionnaires are usually briefer than replies to interview questions: this may not be important where the required answer is of the Yes, No, Don't Know variety, but with more detailed answers it is a disadvantage.

On the whole, questions have to be more carefully worded in a questionnaire than in interviewing, because there is no opportunity of amplifying the point that one is concerned with.

The interview. There are a number of different types of interview, mainly differing in the amount of freedom left to the respondent. Four common types are: (i) *interviewing with a schedule*. A schedule is the interviewer's equivalent of a questionnaire – a detailed list of questions, with space for answers, clipped to a board. The interviewer writes down the answers in the interview itself. The advantages are the same as for the questionnaire, except that there is less time for reflection before answering. There may be some errors of transcription, if the answers are complicated or respondents speak quickly. On the other hand, there is usually a higher percentage of replies from the sample, because there is personal contact and less for the respondent to do. It is sometimes said that interview schedules interfere with rapport, but there seems little evidence to support this view. The Kinsey surveys of one of the most emotionally loaded of all topics – sexual attitudes and behaviour – used interview schedules. Gorer (1955) has an interesting discussion (pp. 313 et seq.) on the relative merits of interview and questionnaire in his survey of personal characteristics commonly found in English men and women. (ii) *partly standardized interviewing*. In this type of interview, a definite set of questions is used, but there is no schedule and consequently a more informal approach is possible. The order of questions can be varied in order to make the interview more 'natural' and to preserve the flow of responses. If some answers seem incomplete, supplementary questions can be asked, and a fuller picture can be gained of the way in which the respondent views the survey topic. Notes can be taken during or after the interview, whichever seems

more appropriate. With this approach, errors of transcription are usually greater than with a schedule, and certain answers may be forgotten altogether. The interview notes tend to reflect the personal style of the interviewer. Some of the material quoted in survey reports shows clear signs of such re-wording by the interviewer (this is particularly noticeable in Zweig's reports, interesting and valuable though they are in many ways.) Sometimes an inexperienced interviewer will omit a number of questions through varying the sequence of the interview. But this type of interview, through its flexibility, can adapt to individual differences between respondents and possibly obtain more detailed and meaningful answers than with the schedule.

(iii) *focused interviewing*. This is a particular kind of partly standardized interview, mainly developed by R. K. Merton (1956). The focus is on individual reactions to an experience that the respondent is known to have had, such as reading a specific book, seeing a film, or taking part in a known social situation. Because the interviewer has made it his business to know about the event that has been experienced, and has carefully analysed it, he is alert to omissions, to idiosyncratic interpretations and to the feelings and thoughts associated with different aspects of the event. (iv) *open-ended or unstructured interviewing*. Here the questions are less specific and the aim is to discover the respondent's personal slant on the topic under consideration. There is general agreement among interviewers that this technique often arouses great interest among respondents and yields more detailed answers, because those being interviewed are in greater control of the situation. But considerable skill is needed on the part of the interviewer if all relevant aspects of the survey topic are to be covered. Interview notes gathered by this means are frequently full and interesting but difficult to compare with one another. Unstructured interviewing can be of very great value in the early stages of an investigation. Errors and omissions can be avoided if a tape-recorder is used (the increasing popularity of these machines, especially for home entertainment, seems to have led to their acceptance in interviewing). But it will be remembered that this gain in accuracy costs a good deal of additional time.

In the present survey, in which we have clearly specified hypo-

theses and a limited time in which to gather relevant information, the interview schedule is the most appropriate technique of research. The first step in drawing up a schedule is to list the factors that have been referred to in the hypotheses. These are: (a) academic achievement, (b) participation in university social life, especially the organized aspects, (c) non-academic achievement in the university, (d) active interest in planning the use of one's time, (e) knowledge of university social activities and aspects of university life other than the course of studies. The second step is to express these as concretely as possible, so that specific information may be collected about them in an interview. Appropriate items would be:

(a) *Academic achievement.* (i) type of course taken (Honours, General, etc.). (ii) examination grades, where these are available. (iii) proportion of passes in annual and terminal examinations (this may be a useful item of information where the information obtained under (ii) is likely to be meagre).

(b) *Participation in university social life.* (i) membership of student societies, including sporting activities. (ii) frequency of attendance at society meetings. (iii) frequency of contacts with staff members. (iv) frequency of use of Union. (v) number of friends inside and outside the course of study. (vi) attendance at lectures outside the course of study. (Note: some thought should be given to the term 'friend'. Perhaps 'anyone who is more than a nodding acquaintance' is a useful meaning in this context).

(c) *Non-academic achievement in the university.* (i) offices held in student societies including Union (e.g. chairman, secretary, treasurer). (ii) membership of university teams (sporting, debating, etc.). (iii) positions in the university associated with high status (e.g. editor of student publication).

(d) *Active interest in planning use of time.* (i) drawing-up and using a timetable of study, (ii) expression of desire not to waste time in the university, (iii) criticism of some university activities as time-wasting.

(e) *Knowledge of university activities, etc.* (i) accurate information about recent events in university life (e.g. important staff changes, new buildings, policy decisions affecting university organization or activities, changes in numbers of students in university). (ii) accurate

information about fundamental aspects of university life (e.g. names of leading staff members, size of university, names of larger student societies).

The third step is to embody these specific points into a series of questions. No attempt will be made here to give an interview schedule suitable for this survey. It is an important part of this practical exercise to draw up one's own schedule, based on local conditions. Here are some guides to making a suitable list of questions. A most important requirement is that questions should be short and plainly worded. In Hamilton Johnston's novel *Dying Nicely*, a young diploma student carrying out a survey for his thesis has a question, 'Do you attach any significance to the relation between housework/breadwinning, churchgoing/drinking, hell-fire/salvation?' and is greatly puzzled by one reply, 'Yes, our Tom and his nippers.' It is better to have a larger number of straightforward questions than a smaller number of 'omnibus' questions. If rather abstract terms are used in the questions, one can expect variations in the replies arising from diverse interpretations of the terms. Questions should follow one another in a logical sequence as far as possible. The actual phrasing should be informal but not 'leading' (that is, they must not suggest that one form of answer is preferable to another).

It is a good idea to start with the easier and more interesting questions and leave more difficult questions until later in the interview. In this survey, for example, more difficulty may be encountered in replying to questions about examination results (particularly where failures are involved) than to questions about social interests and activities. In questions about the frequency with which an activity is undertaken, personal knowledge about the activity is helpful. For example, if one is trying to find out how often a student sees members of the teaching staff for tutorials, discussions, etc., it may be unhelpful to ask, 'How many members of staff have you seen during the last week? What was the occasion on which you met them?' if the last week happens to be one in which staff-student contacts are unlikely anyway. Useful information about drawing-up questions will be found in Moser (1958), Chapters 11 and 12.

Selecting the sample. If two students are conducting the survey, a

total sample of forty-five, including five for a 'pilot survey', will be quite large enough with the limitations of time stated. (Unusually industrious students may manage up to 100.) Since examination performance is an important part of the information to be gathered, the sample should be restricted to second and third years of study. If a larger number of students are available as interviewers, a correspondingly larger sample can be interviewed, and in the analysis of results it will be profitable to divide the sample into appropriate subgroups of faculty, year of study, sex, subject of study, etc.

With the recent development of large-scale surveys, a great deal of attention has been given to sampling. Of the numerous books available, Moser (1958) and Parten (1950) are the most solidly informative, and Madge (1953) the most readable and interesting. Though the terminology of sampling is not standardized, the following summary of approaches is close to current practice. There are two broad divisions – probability and non-probability sampling. In probability sampling, a method of selection is used in which each member of the group to be studied has a calculable probability of being selected (the probability being greater than zero).

With a probability sample it is possible to estimate the likely sampling errors.

Each of the two broad divisions can be divided into three subdivisions. Non-probability sampling can be divided into accidental sampling, quota sampling and purposive sampling. Accidental sampling is that in which the sample is only apparently picked at random; where, for example, a journalist talks to the first four people whom he happens to meet on a busy corner. If one examines the situation in which such sampling takes place, it becomes clear that certain individuals and categories of people are very likely to be passing the corner at the time at which the journalist is interviewing, whereas other individuals and categories are not at all likely to be there. A bias in the interviewing immediately results from the 'accident' not being random. In quota sampling, accidental sampling is made within a number of categories which have been selected as representative of the group being sampled. These usually include age, sex, occupational grouping, region, and other categories for

which accurate population statistics are available. An example of quota sampling would be the selection of a sample of individuals in representative proportion of men and women, age-levels, and occupational groups to the population being sampled. An individual interviewer might then be asked to interview twenty people, of whom ten would be men and ten women, half above the age of forty and half below, one fifth earning more than thirteen hundred pounds a year, and so on. But the selection of individuals to be interviewed would be left to the interviewer, as long as they come within the categories listed. Unless he selected the individuals by a probability method (in which case he would be undertaking stratified random sampling), his selection would be technically 'accidental' and therefore open to various forms of bias. In purposive sampling, the judgement of the research worker is used to select supposedly typical cases from the population being sampled.

Probability sampling, or random sampling, can be divided into simple random sampling, in which selection is made haphazardly (e.g. by means of a table of random numbers) from the group to be sampled; stratified random sampling, in which the group is divided into various parts, or strata (such as age, sex, and occupational groups), within which random sampling is made; and cluster sampling, used in large-scale surveys where obtaining a list of all individuals in the total group would be impossible or impracticable. In cluster sampling, the units which are randomly sampled are not individuals, but clusters of individuals, such as schools, factories or residential areas. When clusters have been randomly selected, they can in turn be sampled randomly, down to the individual level.

This classification of sampling procedures leaves unmentioned the one recommended in this survey – systematic sampling from a list. In this, probably the most widely used of all sampling methods, items or individuals are selected at regular intervals from a list, the first name being picked at random.

The simplest procedure is to obtain a list of students registered at the university in which you are conducting the survey. The names will usually be grouped according to faculties and subjects of study, and the arrangement within each group will be alphabetical (not

strictly random, but a close approximation to it, if the whole list is sampled). If the size of the sample that one is able to manage is divided into the number in the total list, one then knows the sampling interval to use. Thus forty-five names in a population of 4,500 requires every hundredth name to be chosen. The first name should be chosen at random. The fact that the original list is 'stratified' into subjects of study represents a gain in the adequacy of the sample, since the different subject groups will then appear in their correct proportions in the sample. First-year students should not be included in the sample, as insufficient examination data will be available.

Pre-testing: the pilot survey. If five names are chosen randomly from the list of forty-five already obtained, they will serve for testing the adequacy of the interview schedule. The sole purpose of the 'pilot survey' is to gain actual experience of the effectiveness of one's research instruments. With an interview schedule, it will be possible to test the clarity of the questions, the length of time taken by the interview, the sequence of the questions, and the ease with which the schedule can be completed by the interviewer.

Collecting information. It is now possible to begin collecting the survey data. It is seldom feasible to arrange interviews like dental appointments, at half-hourly intervals. Particularly in the university, respondents must be interviewed on their own terms, and this may well mean a good deal of time spent on making provisional arrangements. It is well worth spending some time on such preparations, in order to avoid wasting far more time with interviews spread over a period of weeks.

A tendency which is often found in interviewing is 'interviewer's itch' - the anticipation of what the respondent is going to say. Particularly when a large number of interviews are being undertaken in a short time, one must guard against leading questions, completing the respondent's replies for him, and other demonstrations of impatience which seem to be very elementary errors in cold print but are all too easily made in practice. It will usually be found that respondents have little objection to having their replies written down (often, quite the contrary, an interviewer is expected to make notes as an indication of interest and competence). But if objection should be

made, notes should be written up as soon as possible after the interview, in order to minimize errors and omissions. It is a sound procedure to have interview times arranged so that some time is available between interviews for additional notes, a quick revision, and a breathing space. Continuous interviewing, even over a short period, quickly yields diminishing returns. And it is always wise to make clear to respondents that there will be no personal reference to them in the survey report. (A simpler way of putting it – Nothing you say will be linked up with you in any way. This is absolutely confidential.)

It is at the interviewing phase (and later, when results are being analysed) that pre-coded questions will be found particularly useful. Pre-coding refers to the provision of a limited number of categories covering all appropriate replies to a question (for example, a list of faculties of study; or Yes, No, Don't Know, No Answers). The interviewer then merely ticks the appropriate category on his interview schedule.

Preparing the information. Once the data have been collected, they must be conveniently grouped so that they can be related to the problems from which the survey started. There are many ways of tabulating and classifying data (see especially Sellitz *et al.* (1959) and Zeisel (1958)). In this survey, it is best to transfer the material to 6 × 4 inch index cards, and then to prepare tables which will show the relationships between the variables referred to in the hypotheses.

There are two important points to note in the preparation of data: (i) tables must be readily understood. That is, they must be simple in structure. Two or three tables are preferable to one table with a large and confusing number of columns and rows. They must also clearly indicate the headings under which material is tabulated, and enumerate the number of cases on which the table, or its parts, have been based. It is useful, where large numbers are involved, to use percentages to facilitate comparison. But in very small samples, percentages are usually misleading. In the present survey, one individual constitutes 2.5 per cent. of the total sample (omitting the pilot survey sample). (ii) statistical checks must be appropriate to the techniques of measurement used in the collection of data. Statistical methods based for example on random collection of data, normal distribution

of the characteristics being studied, or equal-interval scales are inappropriate, and give misleadingly precise results, when used on data collected non-randomly, or using material which is only rank-ordered.

The main tests of significance used in simple survey data are the chi-squared test and the t-test. The two main techniques of correlation are the rank-order coefficient (often called the Spearman coefficient) and the product-moment coefficient (or the Pearson coefficient). The formulas for these, and the techniques by which they are used, will be found together with detailed examples in Connolly and Sluckin (1957), Edwards (1946), and McNemar (1949). Techniques which make no assumptions as to the distribution of characteristics in the population from which the sample has been drawn, and are therefore called distribution-free or nonparametric statistics, are described in Siegel (1956) and in Chapter 12 of Festinger and Katz (1954).

The chi-squared test will be found adequate for the present survey data.

Interpretation of the findings. In this phase of the survey, the findings are related to the hypotheses and their implications are discussed. Whatever relationship is found between academic achievement and participation in university life, it will be recalled that many factors have been left uncontrolled and unexamined. Thus, any attempt at giving a simple causal explanation of the relationship would be misplaced. But, as a careful study of some important aspects of university life and work, the survey will be of interest to many, and the implications of the findings should be considered in detail. Particular attention should be given to those findings which surprise the research workers. For it is these that provide the challenge to further investigation.

Presentation of the findings. As always in matters of presentation, the reader of the survey report is the person to be considered. What is it that he wants to know? First, the questions with which the survey started, the itch that prompted the scratch of research. Then he wants to know something of the previous work on these questions. This puts the survey in perspective. Then, the hypotheses, if they are more rigorously stated than the original questions, should be listed. Next, the techniques used in the survey should be described, and the

sample referred to. Then the results should be presented, in the simplest form compatible with accuracy. The significance of these results to the starting questions is then discussed and the survey is finally summarized as briefly as possible. In short, there are six sections of the survey report: (i) what you were trying to do, (ii) what others have done about it, (iii) what was done about it in the survey, (iv) what was discovered, (v) interpretations and reservations, (vi) a summary.

Direct quotations from interview notes, where relevant, can make the survey report much more readable. The survey reports of the Institute of Community Studies offer excellent examples of the use of quotations (for example, Young and Willmott (1957)).

The small-scale survey as a practical exercise

In the first part of this chapter, the values of social surveys have been indicated. What is the value of a student exercise such as we have been working on here? Apart from developing skill in the conduct of a survey, there are many possible advantages: (i) it provides useful practice in reducing ideas to a reasonably precise and testable form, (ii) it increases interest in the studies that are reviewed as relevant to the topic that one has decided to survey, (iii) one increases one's awareness that every piece of research is a compromise between varying demands and limitations, (iv) it sharpens the distinction, sometimes blurred in reading research reports, between findings and interpretation.

The social survey as a collaborative activity

Social research is most profitably conceived as a collaborative activity between research workers and the public. This is clearly seen in the dependence of research workers on the goodwill of those who are to be involved in the project, the concern with rapport, and the interest of the public in the findings of those researches that are related to their interests. Foote and Cottrell (1955) write of 'social science as systematic self-scrutiny'. Of all forms of social and psychological research, the social survey comes closest to truly collaborative activity, since it encounters people in the ordinary and extraordinary

affairs of everyday life and reflects as accurately as possible what is going on. When this knowledge is made available to the widest possible audience, particularly to those whose activities are described, with due regard to the privacies of life, something has been achieved in furthering reasonable conduct. Without vision the people perish, and as society becomes more complicated and volatile, surveys can play a limited but valuable part in sharpening our sense of one another's lives and the ways in which we are bound up with one another.

REFERENCES

- ABRAMS, M. (1951). *Social Surveys and Social Action*. London: Heinemann – a brief and readable account of the use and development of social surveys in Britain and the U.S.
- ABRAMS, M. (1959). *The Teen-Age Consumer*. London: London Press Exchange.
- BIRCH, A. H. (1959). *Small-Town Politics*. London: Oxford Univ. Press – includes an account of a social survey of Glossop, linked with data gained from other sources. Attempts to relate political life to the general social life of the town.
- BELSON, W. A. (1960). *Television and the Family*. London: B.B.C. – a rigorous sample survey of the effects of television-viewing on aspects of family life in Britain.
- CARADOG JONES, D. (1949). *Social Surveys*. London: Hutchinson – gives brief but detailed accounts of British poverty surveys.
- CONNOLLY, T., and SLUCKIN, W. (1957). *Introduction to Statistics for the Social Sciences*. London: Cleaver-Hume – see especially Chapters 8, 9, and 10 on statistical significance, correlation principles and methods.
- EDWARDS, A. L. (1954). *Statistical Methods of the Behavioral Sciences*. New York: Rinehart.
- FESTINGER, L., and KATZ, D. (eds.) (1954). *Research Methods in the Behavioral Sciences*. London: Staples – see especially Chapters 8, 11, and 12 on interviewing, social measurement, and distribution-free statistics.

- FOOTE, N. N., and COTTRELL, L. S. (1955). *Identity and Interpersonal Competence*. Chicago: Univ. of Chicago Press – see especially Chapter I on social science as systematic self-scrutiny.
- GORER, G. (1955). *Exploring English Character*. London: Cresset – an interesting and ingenious attempt to study English national characteristics by means of large-scale survey techniques; see especially Chapters I, 2, and Appendix 2.
- GEDDES, D. P. (ed.) (1954). *An Analysis of the Kinsey Reports*. New York: Mentor – see especially Hyman and Sheatsley on the scientific method.
- HIMMELWEIT, H., OPPENHEIM, A. N., and VINCE, P. (1958). *Television and the Child*. London: Oxford Univ. Press – an elaborate large-scale survey of the effects of television-viewing on the attitudes and activities of children. Interesting integration of qualitative and quantitative data, admirable presentation, carefully-considered recommendations clearly related to findings.
- HIMMELWEIT, H., and SUMMERFIELD, A. (1950/1951). 'Student selection: an experimental investigation, I, II, and III.' *British Journal of Sociology*, I and 2.
- JOHNSTON, HAMILTON (1960). *Dying Nicely*. London: Gollancz – mortifying account, inter alia, of a one-man small-scale survey.
- KUPER, L., et al. (1953). *Living in Towns*. London: Cresset – studies of urban life with frequent use of survey techniques.
- MCNEMAR, Q. (1949). *Psychological Statistics*. New York: Wiley – see especially Chapters I to 8, and II, 12.
- MADGE, J. (1953). *The Tools of Social Science*. London: Longmans – very clear and readable, see especially Chapter 4 on interviewing and Chapter 5 on experiment.
- MALLESON, N. (1958). *Student Performance at University College, London*. Universities Quarterly.
- MERTON, R. K., FISKE, M., and KENDALL, P. L. (1956). *The Focused Interview*. Glencoe, Illinois: Free Press.
- MOGEY, J. M. (1956). *Family and Neighbourhood*. London: Oxford Univ. Press – interesting use of survey techniques in a pilot study of family life in old and new areas of a town.
- MOSER, C. A. (1958). *Survey Methods in Social Investigation*. London: Heinemann – a systematic presentation of problems involved in

the organization of large-scale surveys: much of the material is pertinent to all surveys.

PARTEN, M. (1950). *Surveys, Polls and Samples: Practical Procedures*.

New York: Harper – see especially Chapters 6, 15, and 16.

PEAR, T. H. (1955). *English Social Differences*. London: Allen & Unwin – see especially Chapter 10.

POLANYI, M. (1958). *Personal Knowledge*. London: Routledge – attempts to show that science is an expression and affirmation of personal values, though the search for understanding is pursued with the intention of satisfying the most rigorous and general standards of adequacy.

SELLITZ, C., JAHODA, M., DEUTSCH, M., and COOK, S. W. (1959). *Research Methods in Social Relations*. New York: Holt – a revised one-volume edition of a deservedly standard work: very readable and well-organized, see especially Chapter 5 on measurement and Appendix B on sampling.

SIEGEL, S. (1956). *Nonparametric Statistics*. New York: McGraw-Hill – the first extensive text on statistical methods that make minimal assumptions about the groups from which the samples are drawn.

SIMEY, T. S., and SIMEY, M. B. (1961). *Charles Booth: Social Scientist*. London: Oxford Univ. Press.

WILLMOTT, P., and YOUNG, M. (1960). *Family and Class in a London Suburb*. London: Routledge – extends YOUNG and WILLMOTT (1957) into a London suburb which exemplifies both middle- and working-class ways of life.

YOUNG, M., and WILLMOTT, P. (1957). *Family and Kinship in East London*. London: Routledge – a very readable and well-integrated study of urban and housing-estate family life in the working classes, making resourceful use of interview schedules.

YOUNG, P. V. (1949). *Scientific Social Surveys and Research*. New York: Prentice-Hall.

ZEISEL, H. (1958). *Say it with Figures*. London: Routledge – see especially Pt. 2, on the causal analysis of tabulated data.

ZWEIG, F. (1952). *The British Worker*. London: Penguin – interesting combination of interview material with other data. Lively presentation.

AUTHOR INDEX

- Abrams, M., 183 f.
Addams, J., 184
Adorno, T. W., 35
Allport, G. W., 25, 81
Anthony, E. J., 106
Argyle, M., 18, 106, 108, 124
Aristophanes, 4
Aristotle, 4
Asch, S. E., 24, 34, 45, 156
Atkinson, J. W., 13, 56 f., 65

Bales, R. F., 108, 118, 124
Bartlett, F. C., 81, 84, 136
Bavelas, A., 116 f., 122 f.
Belson, W. A., 188
Berelson, B., 87
Bevan, W., 36
Bion, W. R., 106
Birch, A. H., 182
Bjerkstedt, A., 163, 171
Blake, R., 96
Blum, G. S., 72
Blyth, W. A. L., 163
Booth, C., 184
Borgatta, E. F., 106, 118, 124, 163
Brahmachari, S., 71, 75
Brady, K., 33, 36, 96
Bronfenbrenner, U., 172
Bronowski, J., 6
Brown, R. W., 26 f.
Bruner, J. S., 42, 46 f.
Brunswik, E., 43
Bullimore, J., 24

Cannell, C. F., 131, 133
Cantril, H., 134

Caradog Jones, D., 184
Carlsmith, J. M., 15
Carlyle, T., 3
Cartwright, D., 108, 116
Chapple, E. D., 108
Child, I. L., 15
Clark, R. A., 56, 66 f.
Cohen, J., 43
Conolly, T., 199
Cook, S. W., 198
Cooley, C. H., 107
Cottrell, L. S., 200
Criswell, J. H., 163
Cronbach, L., 57
Crossley, H., 135
Crown, S., 97
Crutchfield, R. S., 16

Darlington, C., 6
Delott, J., 136
Dennis, W., 96
Deutsch, M., 198
Dixon, N. F., 23
Dornbusch, S. M., 44

Edwards, A. L., 199
Einstein, A., 8
Eysenck, H. J., 97

Fahie, J. J., 4
Faust, W. L., 124 f.
Feldman, H., 16
Feshbach, G., 15
Festinger, L., 15, 199
Fillenbaum, S., 38 f.
Finney, D. J., 178

AUTHOR INDEX

- Fiske, M., 192
 Fleming, C. M., 163
 Foote, N. N., 200
 Foulkes, S. H., 106
 Freud, S., 15, 75, 107

 Galileo, 4
 Gardner, R. C., 38 f.
 Geddes, D. P., 186
 Gibb, J. R., 128
 Gilbert, G. M., 36, 98, 100
 Glueck, S., 14, 18
 Goror, G., 191

 Haire, M., 161
 Hare, A. P., 106, 118, 124
 Hartley, E. L., 109, 116
 Hastorf, A. H., 44
 Heise, G. A., 118
 Helmholtz, H., 7 f.
 Herodotus, 183 f.
 Heyns, R. W., 108
 Himmelweit, H. T., 87, 90, 176, 188 f.
 Hochberg, J., 22
 Hodgson, R. C., 38 f.
 Homans, G., 106, 108
 Horan, E., 52
 Howard, J., 183
 Hull, C. L., 17
 Hyman, H. H., 133, 135, 137, 186

 Isaacs, S., 106

 Jahoda, G., 12
 Jahoda, M., 198
 Jenner, E., 7
 Johnson, H., 194
 Judd, J. M., 31

 Kahn, R. L., 131, 133
 Katz, B., 36
 Katz, D., 33, 36, 96, 134, 199

 Katz, L., 171
 Kelley, F., 184
 Kelley, H. H., 109, 124
 Kelvin, Lord, 7
 Kendal, P. L., 132
 Kerr, M., 172
 Kinsey, A. C., 135
 Klein, J., 107
 Kornhauser, A., 133
 Kuper, L., 183

 Lambert, W. E., 38 f.
 Lasswell, H. D., 87
 Latschma, R., 178
 Leavitt, H. J., 116 f.
 Le Bon, G., 107
 Lenneberg, E. H., 25 f.
 Le Play, F., 183
 Levin, H., 16, 75
 Lewin, K., 57, 108
 Lindzey, G., 108, 163
 Lippitt, R., 108
 Lippmann, W., 96
 Loomis, C. P., 163

 Maccoby, E. E., 16, 75, 109, 116, 132
 Maccoby, N., 132
 MacRae, D. G., 155
 MacLeod, R. B., 21
 Mach, E., 8
 Madge, J., 195
 Malleson, N., 189
 Mannheim, H., 18
 Mayhew, T., 184
 McClelland, D. C., 55 f., 62 f.
 McClelland, W., 18
 McCurdy, H. G., 23
 McDougall, W., 1
 McNemar, Q., 199
 Mead, G. H., 107
 Meehl, P. E., 18
 Merton, R. K., 132

AUTHOR INDEX

- Milgram, N. A., 24
 Miller, A., 37
 Miller, G. A., 118
 Miller, N. E., 17
 Mintz, A., 109 f.
 Moge, J. M., 183
 Moreno, J. L., 108, 163
 Morris, J. F., 12
 Moser, C. A., 184 f., 194 f.
 Mowrer, O. H., 17
 Murray, H. A., 58
 Muscio, B., 150

 Nehnevejsa, J., 163
 Newcomb, T. M., 109, 116,
 134
 Newhall, S. M., 31
 Newton, I., 9
 Nickerson, J. F., 31
 Northway, M., 163

 Oppenheim, A. N., 12, 79, 87,
 188
 Osgood, C. E., 57, 71 f.

 Parry, H., 135
 Parten, M., 195
 Pavlov, I. P., 8
 Payne, S. L., 150
 Pettigrew, T. F., 25
 Piaget, J., 106
 Polanyi, M., 181
 Postman, L. J., 81
 Price-Williams, D. R., 12
 Proclus, 3
 Proctor, C. H., 163

 Raven, J. C., 176
 Razran, G., 37
 Richardson, S. A., 44
 Rivers, W. H. R., 24 f.
 Roberts, J. M., 25 f.
 Robinson, P., 18

 Roby, T. B., 56
 Rosenberg, M., 137
 Rowntree, B. S., 184

 Sanford, N., 56
 Schachter, S., 15
 Sears, R. R., 16, 75
 Secord, P. F., 36
 Sellitz, C., 198
 Sensibar, M. R., 67
 Shapiro, D., 46 f.
 Shapiro, M. B., 172
 Sheatsley, P. B., 186
 Siegel, S., 13, 64 f., 175, 199
 Silberman, L., 163
 Simey, M. B., 185
 Simey, T. S., 185
 Simmel, G., 107
 Singer, B. R., 23
 Sluckin, W., 199
 Smythe, D. W., 89 f.
 South, E. B., 124
 Spearman, C. E., 74
 Spice, B., 163
 Spinley, B. M., 172
 Sprott, W. J. H., 106
 Steinzor, B., 108
 Stott, D. H., 14
 Summerfield, A., 189

 Tagiuri, R., 46 f., 171
 Tajfel, H., 23, 35, 41
 Taylor, D. W., 124 f.
 Thibaut, J. W., 109, 124
 Thompson, R. E., 18
 Tinbergen, N., 1
 Toulmin, S., 3

 Veness, T., 1
 Vernon, P. E., 18
 Vince, P., 87

 Welford, A. T., 17

SUBJECT INDEX

- | | |
|--|--|
| <p>White, R. K., 87, 90
 Whyte, W. F., 106, 115
 Wilcoxon, F., 64, 75
 Wilkins, L. T., 18
 Wilmott, P., 183, 187, 200
 Wilson, T. R., 171
 Wishner, J., 45 f.
 Wolfenstein, M., 87</p> | <p>Woodward, J. L., 136
 Young, M., 183, 187, 200
 Young, P., 184
 Zander, A., 108, 116
 Zeisel, H., 198
 Zweig, F., 192</p> |
|--|--|

SUBJECT INDEX

- | | |
|--|---|
| <p>achievement, academic, 189 f.
 achievement motivation, 55 f.
 affiliation motivation, 55
 aggression, 16
 alcohol, 67 f.
 anal character, 15
 arousal of motives, 55 f.
 Blacky test, 72
 case-studies, 16 f.
 centrality, 116
 child-rearing, 171 f.
 class, social, 155 f., 134
 coding, 26 f.
 cognitive dissonance, 15
 colour recognition, 26 f.
 communication, 79 f.
 communication networks, 115 f.
 competition, 109 f.
 conflict, 55 f.
 conformity, 24
 content analysis, 86 f.
 co-operation, 109 f.
 delinquency, 14, 18
 emotional contagion, 109 f.
 emotional involvement, 35</p> | <p>emotional significance,
 and perception, 22 f.
 ethics, professional, 186
 experiments, 3 f.
 fantasy, 15
 friendship choices, 163 f.
 groups, problem-solving, 109 f.
 group size, 123 f.
 guilt, 71 f.
 heroes, fictional, 90
 hunger, and fantasy, 56
 hypotheses, 13 f.
 identification, 18
 illusions, 24 f.
 individuals, 17
 informants, 172 f.
 interaction recording, 13
 interviewer, role of, 173
 interviews, 131 f, 173 f., 191 f.
 interviews, error in, 134 f.
 introjection, 18
 labels, 32 f.
 language, 25 f.
 laws of social behaviour, 14, 108</p> |
|--|---|

SUBJECT INDEX

- leadership, 110 f.
- learning theory, 17 f.
- level of aspiration, 56, 58 f.

- market research, 182 f.
- mass media, 87
- mass phenomena, 109 f.
- motivation, 55 f.
- mutual choices, 167 f.

- names of colours, 26 f.
- negroes, perception of, 36

- panic, 112 f.
- perception, social, 20 f.
- perception, and social class, 150 f.
- personality inventory, 158 f.
- personality, judgement of, 38 f.
- personality traits, inferring, 42 f.
- philosophy, 1 f.
- practical classes, 11 f.
- prediction, 14, 18
- projection tests, 55 f.
- psychoanalysis, 15, 17

- racial stereotypes, 21, 33 f., 96 f.
- radio programmes, 89 f.
- rapport, 132 f.
- rumour, 80 f.

- sampling, 194 f.

- satisfaction, 119
- semantic differential, 72 f.
- serial reproduction, 81 f.
- sex of interviewer and
 respondent, 137, 148 f.
- sexual behaviour, 135
- sexual motivation, 66 f.
- sexual symbolism, 68 f.
- small social groups, 105 f.
- social surveys, 12, 181 f.
- social surveys, history of, 183 f.
- socialization, 171 f.
- sociomatrix, 167 f.
- sociometry, 163 f.
- statistical methods, 13, 32, 50, 62,
 69, 74, 91, 98, 114, 121, 126,
 148, 152, 160, 170, 178, 198 f.
- stereotypes, 33 f., 96 f.
- subjects, 11 f.
- super-ego, 71 f.
- symbolism, sexual, 67 f.

- T.A.T., 56 f.
- television, 89
- theory, 17

- unconscious, 75

- values, 23, 90 f.
- vices, 139 f.
- voting, 182

